

中国研究生创新实践系列大赛  
“华为杯”第十七届中国研究生  
数学建模竞赛

学 校 北京工业大学

---

参赛队号 20100050019

---

1. 涂闯

---

队员姓名 2. 陈健

---

3. 孔丽

---

中国研究生创新实践系列大赛

# “华为杯”第十七届中国研究生 数学建模竞赛

## 题 目 汽油精制过程中的辛烷值损失预测及优化研究

### 摘 要：

为减轻汽车尾气排放对世界生态环境的压力，国内外环保组织对车用汽油硫含量提出了日趋严格的要求。根据我国环保部发布的最新规定，从 2020 年 7 月 1 日起，要求所有销售、注册登记的汽车必须符合国 VI (a 标准)，从 2023 年 7 月 1 日起，要求所有销售、注册登记的汽车必须符合国 VI (b 标准)，并长远要求汽油指标符合国 VI 标准，使得低硫、超低硫将继续成为未来车用汽油的发展趋势。因此，采用高效的技术手段降低催化裂化汽油中的硫含量成为目前汽油生产过程的重中之重。

S-Zorb 装置脱硫技术具有脱硫率高、辛烷值损失小、成本低等优点，可以在得到较高脱硫率的同时保证较小的辛烷值损失，被国际上称为有生命力的新工艺技术。本文针对 S-Zorb 装置控制系统变量庞杂的问题，通过运用灰色关联度分析明晰其中关键影响指标，对 S-Zorb 装置的重点设备进行控制，从而优化 S-Zorb 装置，充分激发 S-Zorb 装置的最大效用潜能。同时基于这些重点变量建立神经网络及遗传算法，厘清主要操作变量优化调整过程中对应的汽油辛烷值和硫含量的变化轨迹。

问题一：对异常数据的检验与剔除是保证原始数据可靠性与研究准确性的前提。基于此需求，本问题依次采用最大最小的限幅剔除异常值法、拉依达准则 ( $3\sigma$ ) 准则对样本数据中的异常数据进行检测，进而根据研究要求进行数据删除与补齐。

问题二：S-Zorb 装置脱硫过程涉及数量庞大的操作步骤及相应变量，变量之间具有高度非线性和相互耦联关系。因此，本问题选用适用于此问题的灰色关联度分析对关键变量进行筛选。以产成品的辛烷值含量为自变量，分别验证其他变量与产成品间的相关度，提取相关度大于 80% 的变量（共计 30 个），作为后续研究的要素指标。结果涵盖 2 个原材料属性变量（辛烷值、饱和烃, %）和 28 个操作变量。

问题三：考虑到脱硫过程实则是一个非线性复杂系统，本研究以筛选出的 30 个变量指标作为输入层数据、以产成品的辛烷值和含硫量作为 2 个输出层数据指标，构建了加入了动量因子的三层 BP 神经网络预测模型，并对该网络的隐含层数、学习率等关键指标进行对比分析及网络拟合度检测，结果表明构建的神经网络拟合度较好。

问题四：针对产成品辛烷值损失幅度小于 30% 和含硫量低于  $5\mu\text{g/g}$  的优化要求，本研究以 30 个变量的取值范围、产成品辛烷值损失幅度小于 30% 和含硫量小于  $5\mu\text{g/g}$  为约束条件，以产成品辛烷值损失最少为目标，构建优化模型，并运用遗传算法对优化模型进行求解。在算法设计中将构建的 BP 神经网络预测模型嵌入遗传算法中以获取 28 个操作变量值变化（2 个原材料属性变量不变）对应的产成品的硫含量和辛烷值数值。

问题五：在筛选出的 28 个关键操作变量的取值范围内，分别从最小值开始，以各变量给定的变化值为步长，代入构建的产成品辛烷值损失预测模型，观测产成品的辛烷值和含硫量的变化，并分别绘制曲

线图进行可视化展现，进而讨论各操作变量变化对产成品硫含量和辛烷值的影响。

**关键词：**汽油精制过程；辛烷值预测；优化；灰色关联；BP 神经网络；遗传算法

## 目录

一、问题重述 .....	4
1.1 问题背景.....	4
1.2 问题提出.....	5
二、问题分析与求解.....	5
2.1 数据处理.....	5
2.1.1 采用最大最小的限幅方法剔除异常值.....	6
2.1.2 依照拉依达准则 ( $3\sigma$ ) 准则去除异常值.....	6
2.1.3 样本剔除及代替.....	6
2.2 模型主要变量筛选.....	7
2.2.1 无量纲化.....	8
2.2.2 灰色关联度计算.....	8
2.2.3 计算结果及变量筛选.....	8
2.3 辛烷值损失预测模型构建.....	10
2.3.1 模型建立.....	11
2.3.2 模型验证.....	12
2.4 辛烷值损失优化模型构建.....	14
2.4.1 模型建立.....	14
2.4.2 模型求解算法设计.....	15
2.4.3 遗传参数确定.....	17
2.4.4 优化操作.....	19
2.5 模型可视化展示.....	22
2.5.1 数据指标处理.....	22
2.5.2 数据预测仿真可视化.....	23
三、模型评价 .....	29
3.1 模型优点.....	29
3.2 模型缺点.....	29
参考文献 .....	30
附录 Matlab 源程序 .....	31
问题一算法程序.....	31
问题二算法程序.....	33
问题三算法程序.....	35
问题四算法程序.....	36
问题五算法程序.....	40

# 一、问题重述

## 1.1 问题背景

为控制汽车尾气排放造成的大气污染，我国加快了汽油质量升级的步伐，汽油向低硫/超低硫和低烯烃、芳烃含量方向发展。我国在 2014 年全面实施国 IV 汽油标准的基础上，将国 V 汽油标准的执行时间由原定的 2018 年 1 月 1 日提前至 2017 年 1 月 1 日，并将分别于 2019 年 1 月 1 日和 2023 年 1 月 1 日起执行国 VI 和国 VIB 汽油质量标准<sup>[1]</sup>。车用汽柴油新标准的发布实施，在减少机动车污染物排放、改善空气质量的同时，也对炼油行业带来了前所未有的挑战。为了有效利用重油资源，我国大力发展了以催化裂化（FCC）为核心的重油轻质化工艺技术，将重油转化为汽油、柴油和低碳烯烃，目前超过 70% 的汽油是由催化裂化生产得到。其中，汽油清洁化的重点是降低汽油中的硫、烯烃含量，同时尽量保持其辛烷值。

辛烷值是车用汽油最重要的指标，汽油燃料的辛烷值是一个国家炼制水平和车辆设计制造水平的综合反应，也是炼油厂生产车用汽油、区别牌号、控制质量、指导生产、调和工艺的重要指标<sup>[2]</sup>。所以，提高汽油的辛烷值并达到相应的标准要求具有重要的现实意义。而现阶段，大力推进脱硫率提高的同时也带来了辛烷值损失的增加。辛烷值损失的直接影响因素是烯烃加氢饱和程度，辛烷值损失会随着烯烃损失的增加而增大，所以减少辛烷值损失的关键是降低脱硫反应过程中的烯烃加氢饱和反应。

汽油吸附脱硫技术（S-Zorb）技术是 ConocoPhillips 公司开发的主要针对 FCC 汽油馏分的吸附脱硫技术，具有脱硫效率高和辛烷值损失小的特点<sup>[3]</sup>。对于 S-Zorb 装置技术，在生产低硫汽油的同时尽可能地降低辛烷值损失是一个最主要的操作目标，也是其主要技术优势之一，是目前工业化应用较广的吸附脱硫技术。截至 2018 年底，我国建成 S-Zorb 催化汽油吸附脱硫装置 35 套，加工了超过 50% 的催化裂化汽油，在油品质量升级过程中发挥了至关重要的作用<sup>[4]</sup>。但在实际运行过程中，S-Zorb 装置的反应过程硫吸附速率受到吸附剂的数量、反应器的压力、氢分压、温度、硫的吸附量等因素的影响；反应过程烯烃加氢反应受氢分压、温度等因素的影响；反应过程还原反应速率受氢气浓度、吸附剂数量等因素的影响；再生过程吸附剂的氧化和活性则受到氧气流量、水蒸气分压等因素的影响。可见，该工艺技术影响因素众多。

综合来看，S-Zorb 装置的反应运行受到众多操作变量的影响和限制，想要充分发挥及挖掘 S-Zorb 装置最大的技术优势，有效降低催化裂化汽油中的硫含量，并使其中的辛烷值损失较小，就需要对各种操作变量进行逐一的调整和克服。但是数量庞大的操作变量之间具有高度非线性和互强耦联关系，这既增加了工程师的工作量，又会造成无法控制重点操作变量的情况出现，使 S-zorb 装置的过程优化响应不及时，大大增加了企业的成本付出。因此，需要充分利用数据挖掘技术，厘清 S-zorb 装置中不同操作变量之间的相互关系，从中寻找出主要影响 S-Zorb 装置反应的重点操作变量，有的放矢的进行全面

系统优化，发挥 S-Zorb 装置的最大使用潜能。在保证充分脱硫的情况下，尽量降低汽油辛烷值损失。

## 1.2 问题提出

为实现在保证脱硫效果的前提下，尽量降低汽油辛烷值损失，本文依次解决如下问题：

### 问题 1：数据处理

参考附件 1 近四年的工业数据的预处理结果，依附件 2 对 285 号和 313 号数据样本进行预处理并将处理后的数据分别加入到附件 1 中相应的样本号中。

### 问题 2：寻找建模主要变量

建立降低辛烷值损失模型涉及包括 7 个原料性质、2 个待生吸附剂性质、2 个再生吸附剂性质、2 个产品性质等变量以及另外 354 个操作变量（共计 367 个变量），工程技术应用中经常使用先降维后建模的方法，这有利于忽略次要因素，发现并分析影响模型的主要变量与因素。因此，需要根据见附件 1 提供的 325 个样本数据，通过降维的方法从 367 个操作变量中筛选出建模主要变量，使之尽可能具有代表性、独立性（为了工程应用方便，建议降维后的主要变量在 30 个以下），并详细说明建模主要变量的筛选过程及其合理性。

### 问题 3：建立辛烷值损失预测模型

采用问题二中样本和建模主要变量，通过数据挖掘技术建立辛烷值（RON）损失预测模型，并进行模型验证。

### 问题 4：主要变量操作方案的优化

在保证产品硫含量不大于  $5 \mu\text{g/g}$  的前提下，利用模型获得附件 4 中 325 个数据样本中，辛烷值（RON）损失降幅大于 30% 的样本对应的主要变量优化后的操作条件（优化过程中原料、待生吸附剂、再生吸附剂的性质保持不变，以它们在样本中的数据为准）。

### 问题 5：模型的可视化展示

工业装置为了平稳生产，优化后的主要操作变量（即：问题 2 中的主要变量）往往只能逐步调整到位，对 133 号样本（原料性质、待生吸附剂和再生吸附剂的性质数据保持不变，以样本中的数据为准），以图形展示其主要操作变量优化调整过程中对应的汽油辛烷值和硫含量的变化轨迹。

## 二、问题分析与求解

### 2.1 数据处理

在实验中不可避免会存在一些异常数据，而异常数据的存在对研究对象的变化规律和发展趋势会产生重要的影响，对数据进行异常值检验是保证原始数据可靠性的前提。多种检验方法同时使用可以提高检测精度，进而提高样本数据的准确性。因此，本文根据相关研究数据问题，依附件二中的样本确定方法，对 285 号和 313 号样本进行数据预处理，主要分为以下三个流程：

### 2.1.1 最大最小的限幅方法剔除异常值

#### step1: 数据读取并找出最大最小值

读取附件四中操作变量范围数据，拆分成 354 个操作变量的最大值和最小值矩阵，为矩阵 **cmax**、**cmin**，均为 1 行 354 列矩阵。然后，分别读取附件三中 285 号和 313 号数据，设置矩阵为 **D285** ( $40 \times 354$ )、**D313** ( $40 \times 354$ )。

#### step2: 数据范围剔除异常值

按照式 (1) 对各数据值进行判定。当 D285 和 D313 中两个矩阵的元素超出各最小值和最大值的区间范围时，设置其值为 0；否则为原值。式中 **X** 为 D285 或 D313 矩阵，*i, j* 分别表示矩阵的行列号。剔除异常之后原有的 D285 和 D313 矩阵形成的新矩阵分别为 LD285 和 LD313。

$$\begin{cases} X(i, j) = X(i, j) & (c \min(j) < X(i, j) < c \max(j)) \\ X(i, j) = 0 & (\text{其他}) \end{cases} \quad (1)$$

### 2.1.2 拉依达准则 ( $3\sigma$ ) 准则去除异常值

首先分别计算 LD285 和 LD313 两个矩阵的算数平均值  $\overline{X(j)}$ ；然后按  $V_{ij} = x(i, j) - \overline{X(j)}$  计算两矩阵各元素的剩余误差；最后基于贝塞尔公式计算出各列标准误差  $\sigma(j)$ ，见式 (2)。其中，与平均值超过三倍标准差的测定值，称为高度异常的异常值，即当  $|V_{ij}| = |x(i, j) - \overline{X(j)}| > 3\sigma$  时，认定为异常值，设置  $x(i, j)$  为 0，否则保持原值。

$$\sigma(j) = \left[ \frac{1}{n-1} \sum_{i=1}^n v_{ij}^2 \right]^{1/2} = \left\{ \left[ \sum_{i=1}^n x_{ij}^2 - \left( \sum_{i=1}^n x_{ij} \right)^2 / n \right] / (n-1) \right\}^{1/2} \quad (2)$$

### 2.1.3 样本剔除及代替

在完成上述步骤数据操作后，查找每行为空值 (0) 的数量。本文剔除原则为若每行残缺数据超过 20 (即每行 '0' 的个数超过 20 个)，则认为数据残缺较多，无法补充，则将此类位点删除，即删除该时间下的整行数据；替代原则为若部分数据为空值，则空值处使用其前后两个小时数据的平均值补齐。基于上述异常值检验结果，经计算 285 号和 313 号样本中，每个时间位点中异常值的数量如下表 1 所示。

表 1 285 号和 313 号数据样本中每个时间位点对应变量的异常值数量

样本编号	285 号	313 号	样本编号	285 号	313 号
1	15	26	21	16	20
2	14	23	22	16	21
3	14	24	23	16	16
4	24	34	24	16	18
5	16	21	25	16	20
6	16	31	26	16	22
7	16	27	27	16	20
8	20	31	28	16	26
9	19	28	29	16	18
10	19	32	30	16	18
11	21	31	31	16	16
12	21	32	32	16	17
13	18	33	33	16	20
14	19	28	34	16	15
15	15	25	35	16	19
16	16	21	36	16	18
17	19	23	37	16	20
18	17	30	38	16	30
19	22	31	39	16	19
20	25	32	40	16	18

根据上表，在 285 号数据样本中，每个时间位点的异常值数量均较少，因此根据数据补齐原则，将残缺或异常数据进行补齐。在 313 号数据样本中，共有 9 个时间位点样本数据残缺大于 20，将此 9 个位点删除。基于上述调整，得到最终样本数据见附件 1-1，然后将各操作变量样本数据的平均值进行计算，结果见附件 1-2，并将该数据处理结果添加到所给附件 1 中。

## 2.2 模型主要变量筛选

利用 S-Zorb 装置对重油进行催化裂化，工艺复杂，设备多样，其工艺流程如图 1 去所示。发挥 S-Zorb 装置技术的最大使用潜力需要对其脱硫过程进行充分优化。但在实际操作中，S-Zorb 装置由若干影响因素构成的，其系统的形成、运行、发展的趋势和规律均由其构成因素来决定。因此，S-Zorb 装置脱硫过程涉及数量庞大的操作变量，变量之间具有高度非线性和相互耦联关系。并且在进行样本数据测试、汇总的过程中，出现了某些样本数据不易获取、大量数据缺失等极端情况。这就需要选择实用性强、易操作、效果显著、能充分利用、挖掘已有数据的数学模型进行辅助研究。而灰色关联分析模型要解决的问题就是部分数据已知，部分信息未知或者缺失的情况，而且该模型克服了优化方案受决策者对相似性操作变量和相近性操作变量的主观偏好影响，通过对已知样本数据的生成和挖掘，筛选出其中最为有价值的数据信息，实现对 S-Zorb 等类似不确定性系统的性质、结构、演化规律的正确描述及有效监控，从而达到优化流程、充分发挥装置潜能的目的。



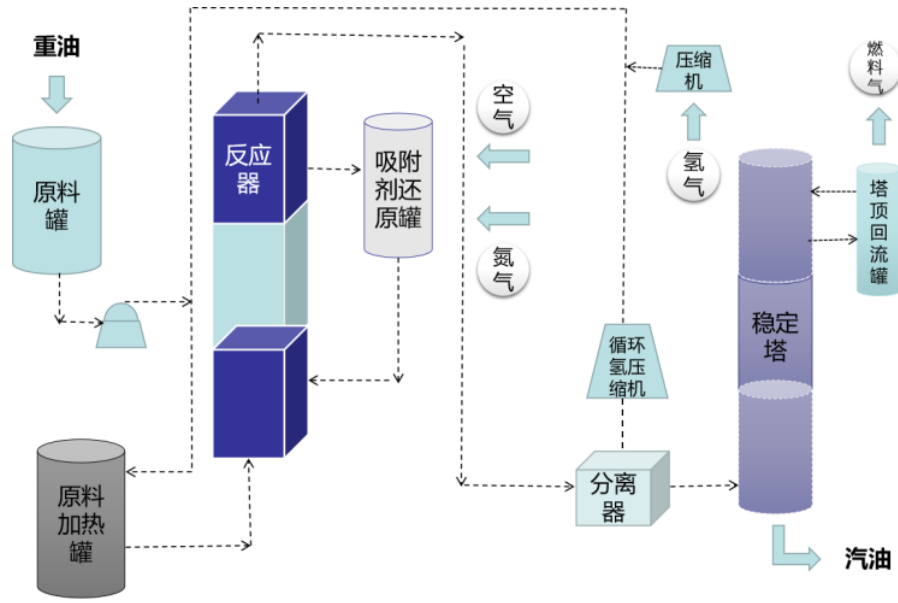


图 1 S-Zorb 装置催化裂化工艺流程

基于上述分析，本文选择灰色关联度方法对数据进行降维与数据筛选。灰色关联度方法基于灰色理论<sup>[5]</sup>，研究事物之间相互关联、相互作用的复杂因素的影响作用，并确定影响事物的本质因素，使各种影响因素之间的“灰色”关系清晰化的分析方法。灰色关联分析的基本思想是通过确定参考数据列和若干个比较数据列的几何形状相似程度来判断其联系是否紧密，它反映了曲线间的关联程度。通常可以运用此方法来分析各个因素对于结果的影响程度，也可以运用此方法解决随时间变化的综合评价类问题，其核心是按照一定规则确立随时间变化的参考序列，把各个评估对象随时间的变化作为因素序列，求各个因素序列与参考序列的相关度，依照相关性大小得出结论。具体计算步骤如下：

#### 2.2.1 无量纲化

由于各指标属性不同，计量单位不同，为了建立统一的评价标准，需要对其进行规范化处理。见式 (3)

$$X_i(k) = \frac{X_i(k)}{\bar{X}_i} \quad (3)$$

其中， $X_i(k)$ 表示第  $k$  个时间点第  $i$  个变量的样本数据值。

#### 2.2.2 灰色关联度计算

令产品性质中辛烷值含量为参考序列  $X_0 = x_0(k), k = 1, 2, \dots, n$ ，因素序列  $X_i = x_i(k), k = 1, 2, \dots, n, i = 1, 2, \dots, m$ ，则  $X_0$  和  $X_i$  的灰色关联度  $r(X_0, X_i)$  定义为：

$$r(X_0, X_i) = \frac{1}{n} \sum_{k=1}^n r(X_0(k), X_i(k)) \quad (4)$$

$$r(X_0(k), X_i(k)) = \frac{\min_i \min_k |X_0(k) - X_i(k)| + \rho \max_i \max_k |X_0(k) - X_i(k)|}{|X_0(k) - X_i(k)| + \rho \max_i \max_k |X_0(k) - X_i(k)|} \quad (5)$$

其中， $\rho$  为分辨系数，且  $\rho \in [0, 1]$ ，其作用在于提高关联系数之间的差异显著性。通常， $\rho = 0.5$ 。

#### 2.2.3 计算结果及变量筛选

根据灰色关联度模型，基于 MATLAB 软件，对各变量与辛烷值的相关度进行计算。便于后期优化

方案的设计，本文筛选相关性较高的 30 个指标变量（相关度大于 80%）。筛选变量结果及相应的相关度见表 2。

表 2 主要变量筛选结果

变量名称	符号表示	相关度
K-102A 进气温度	X <sub>1</sub>	0.991
F-101 出口支管#4 温度	X <sub>2</sub>	0.987
E-101D 壳程出口管温度	X <sub>3</sub>	0.959
干气出装置温度	X <sub>4</sub>	0.951
3#催化汽油进装置流量	X <sub>5</sub>	0.948
<b>*辛烷值 RON</b>	<b>X<sub>6</sub></b>	<b>0.945</b>
混氢点氢气流量	X <sub>7</sub>	0.944
<b>*饱和烃,v%（烷烃+环烷烃）</b>	<b>X<sub>8</sub></b>	<b>0.934</b>
低压热氮气压力	X <sub>9</sub>	0.915
E-101C 管程出口管温度	X <sub>10</sub>	0.893
净化风进装置压力	X <sub>11</sub>	0.892
1.0MPa 蒸汽进装置温度	X <sub>12</sub>	0.889
F-101 出口支管#1 温度	X <sub>13</sub>	0.873
F-101 出口支管#3 温度	X <sub>14</sub>	0.870
稳定塔顶压力	X <sub>15</sub>	0.870
D-113 压力	X <sub>16</sub>	0.867
还原器压力	X <sub>17</sub>	0.864
反应器入口温度	X <sub>18</sub>	0.854
R-102 床层吸附剂料位密度	X <sub>19</sub>	0.854
非净化风进装置压力	X <sub>20</sub>	0.846
再生器下部温度	X <sub>21</sub>	0.844
D-122 液位液位	X <sub>22</sub>	0.843
D-124 液位	X <sub>23</sub>	0.842
R-102 #1 通风挡板温度	X <sub>24</sub>	0.826
再生器下部温度	X <sub>25</sub>	0.821
再生器温度	X <sub>26</sub>	0.819
加热炉排烟温度	X <sub>27</sub>	0.817
F-101 出口总管压力	X <sub>28</sub>	0.817
D104 压力	X <sub>29</sub>	0.806
蒸汽进装置压力	X <sub>30</sub>	0.803

\*表示原料性质变量

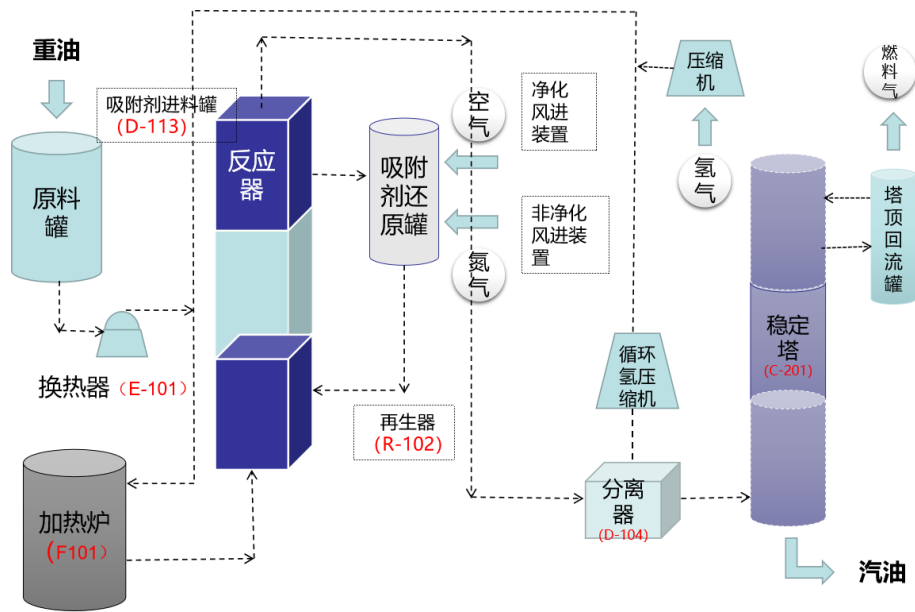


图 2 S-Zorb 装置中重点优化的设备示意图

如表 2 所示，即为利用灰色关联度分析模型进行筛选后的主要操作变量。结合筛选出的操作变量，经过对 S-Zorb 装置各关键设备分析可知，应重点关注 S-Zorb 装置中的加热炉（F-101）、吸附产物换热器（E-101）、吸附剂进料罐（D-113）、再生器（R-102）、缓冲罐（D-124）、气液分离罐（D-104）、稳定塔（C-201）等（见图 2），并且主要对上述设备所涉及到的 30 个重点变量进行优化设置，从而使 S-Zorb 装置达到最佳工作状态。

### 2.3 辛烷值损失预测模型构建

依据汽油加工工艺过程，原材料经过系列加工得到产成品。而产成品的质量主要通过辛烷值、含硫量和辛烷值损失三个主要指标对其描述。其中**辛烷值损失 = 原材料辛烷值 - 产成品辛烷值**。对于石油脱硫加工各个环节，组成了一个非线性复杂系统，各操作环节间的相互作用、配合决定了产品的质量。因此，研究辛烷损失模型要从非线性视角构建，且应以产成品的辛烷值含量和含硫量作为因变量，生产环节操作变量和原材料等要素作为自变量，构建辛烷值损失预测模型。

预测方法按照类型分为定量分析法与定性分析法。定性分析法主要针对预测的资料数据缺失情况，依靠预测人员的丰富经验及主观判断能力来分析。定量分析法是通过数学模型计算出分析对象的各项数值指标，主要是时间序列分析法、因果关系分析法等。神经网络是通过训练而模拟被测系统任意复杂的非线性关系，故其在智能控制、模型预测、模式识别等重大领域具有重要应用。BP 神经网络凭借其良好的函数逼近能力在数据预测领域有大量的研究。例如，Zong<sup>[7]</sup>基于神经网络预测了韩国的交通部门能源需求量，并指出神经网络模型优于线性回归模型和指数模型。王珏<sup>[8]</sup>建立了基于小波神经网络的我国能源需求非线性预测模型。吕彬<sup>[9]</sup>运用 BP 神经网络对电力负荷进行了预测研究。Sözen A<sup>[10]</sup>运用神经网络预测土耳其各部门的能源消费量，后来该神经网络模型被应用到温室气体预测和预警中去。Kankal<sup>[11]</sup>运用神经网络和回归分析的算法和国内生产总值、人口、进出口等社会经济和人口统计数据指标预测土耳其能源消费，并使用相对误差和标准误差（RMSE）验证模型的有效性。除此之外还有许多学者对神经网络算法进行改进，研究预测问题。陶阳威<sup>[12]</sup>等用在误差反向传播时加入动量项对传统的 BP 神经网络进行改进，并以国民生产总值、城镇人口比例、产业结构变化、能源价格、能源结构、技术进步、消费水平等 8 项数据指标作为输入层数据，建立了中国能源需求模型，对中国未来 3 年的能源需求进行了预测。张俊深<sup>[13]</sup>基于对 GM(1,1)模型进行无偏模型改进，并将各种非线性参变量间的映射纳入组合 BP 神经网络的能源消费预测。Beccali M<sup>[14]</sup>使用无监督学习和有监督学习的神经网络的联合方

法预测城市每日电力负荷状况。因此，本研究采用上述样本和建模主要变量，本研究选择 BP 神经网络建立辛烷值（RON）损失预测模型。

BP 神经网络属于前向神经网络，强调网络采用误差反向传播的学习算法。其包括一个输入层、若干隐含层和一个输出层组成。其核心思想是通过样本训练集，不断修正神经网络的权值和阈值，逐步逼近期望输出值。在训练开始时沿着网络正向传播，然后根据网络输出值与期望输出值之间误差，反向传播调整权值和阈值。通过反复更新网络权值和阈值实现误差最小即完成网络训练。

其中输出层指标为问题 2 中筛选出的 30 个变量（见表 2），输出层指标为产成品的辛烷值和含硫量两个指标。然后运用原材料的辛烷值减去模拟出的产成品的辛烷值，即可得到辛烷值损失值。构建的 3 层 BP 神经网络如图 3 所示。

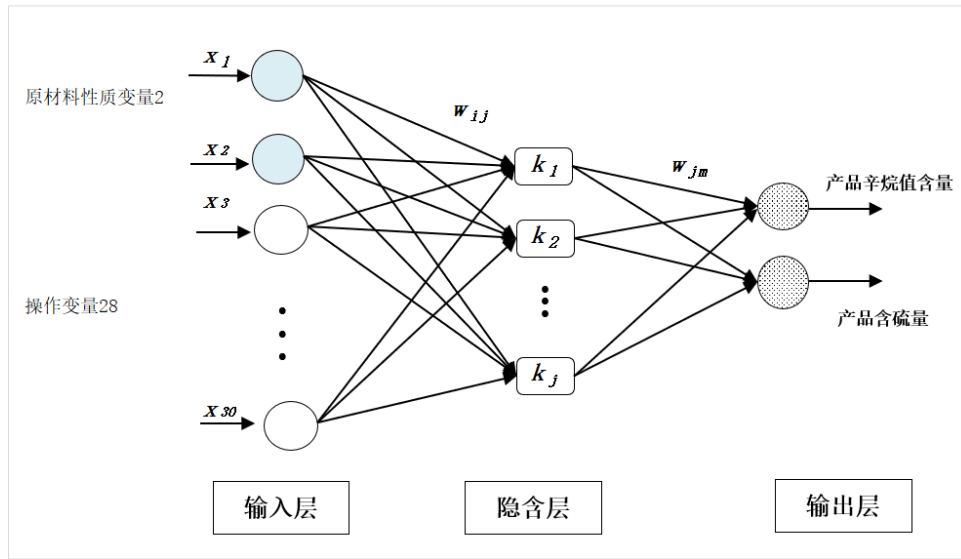


图 3 构建的神经网络结构图

### 2.3.1 模型建立

（1）隐含层与输出层计算公式如下：

① 隐含层第  $j$  个神经元输入值为：

$$hid_{input_j} = \sum_{i=1}^I w_{ij} \cdot x_i \quad (6)$$

② 隐含层第  $j$  个神经元输出值为：

$$hid_{output_j} = F(hid_{input_j}) \quad (7)$$

$F(x)$  为传递函数，由于本文数据均大于 0，传递函数选取为：

$$F(x) = \frac{1}{1+e^{-x}} \quad (8)$$

③ 输出层输入计算公式 为：

$$out_{input_m} = \sum_{j=1}^J w_{jm} \cdot hid_{output_j} \quad (9)$$

④ 输出层计算公式为：

$$out_{output_1} = F(out_{input_1}) \quad (10)$$

（2）计算训练误差

① 第  $n$  代误差为  $em(n)$ ， $y_m$  为期望输出，其中：

$$e_m(n) = out\_output_1 - y_m \quad (11)$$

② 计算网络总误差为:

$$e(n) = \frac{1}{2} \sum_{m=1}^M e_m^2(n) \quad (12)$$

(3) 误差反向传播

① 输出层与隐含层之间的权值修正量为:

$$\Delta w_{jm}(n) = -\eta \frac{\partial e(n)}{\partial w_{jm}(n)} \quad (13)$$

修正后的权值为:

$$w_{jm}(n+1) = w_{jm}(n) + \Delta w_{jm}(n) \quad (14)$$

其中,  $\eta$  为学习率。

② 输入层与隐含层之间的权值调整为:

$$w_{ij}(n+1) = -\eta \frac{\partial e(n)}{\partial w_{ij}(n)} + w_{ij}(n) \quad (15)$$

(4) 动量 BP 法

由于 BP 神经网络存在收敛速度慢的缺点, 针对该不足本文采用动量 BP 法更新权值, 即该更新权值时引入动量因子  $\alpha$  ( $0 < \alpha < 1$ ), 则权值修正公式为:

$$\Delta w(n) = -\eta(1 - \alpha)\nabla e(n) + \alpha\Delta w(n-1) \quad (16)$$

### 2.3.2 模型验证

(1) 训练数据和测试数据确定

通过问题二的指标筛选, 得到了 325 行 30 列的变量数据, 即神经网络的输入层数据, 输出层数据为 325 行 2 列的数据。本研究设置前 300 行数据为训练数据, 后 25 行数据为测试数据, 即在筛选出 30 个变量指标下, 对于附件一中的数据, 以编号 1-300 号的数据为神经网络训练数据, 301-325 号的数据为测试数据。基于此设定, 按照公式 (17) 进行数据标准化。

$$x^* = \frac{x - \max}{\max - \min} \quad (17)$$

(2) 参数设定

本研究设计的神经网络图涉及三个参数, 分别为迭代次数、训练目标、隐含层元素个数和学习率。由于涉及数据较少, 本研究设置迭代次数为 1000, 训练误差目标为 0.000001。然后讨论隐含层元素个数可和学习率的设定。首先设学习率为 0.1, 分别设定隐含层元素个数为: 45, 60, 90, 120。在此基础上讨论不同隐含层元素个数的测试误差。通过对比模拟值和真实值间的误差确定隐含层元素个数, 其中误差计算公式为式 (18)。

$$err(j) = \frac{(\text{含硫量模拟值}(j) - \text{含硫量实际值}(j)) + (\text{辛烷值模拟值}(j) - \text{辛烷值实际值}(j))}{2} \quad (18)$$

不同隐含层元素个数各测试点误差曲线图见图 4, 误差算数平均值见表 3。

表 3 不同隐含层和学习率的误差均值

隐含层元素个数	误差均值	学习率	误差均值
45 个	2.01	0.1	2.01
60 个	2.75	0.4	1.62
90 个	1.74	0.7	2.11
120 个	3.09	0.9	1.81

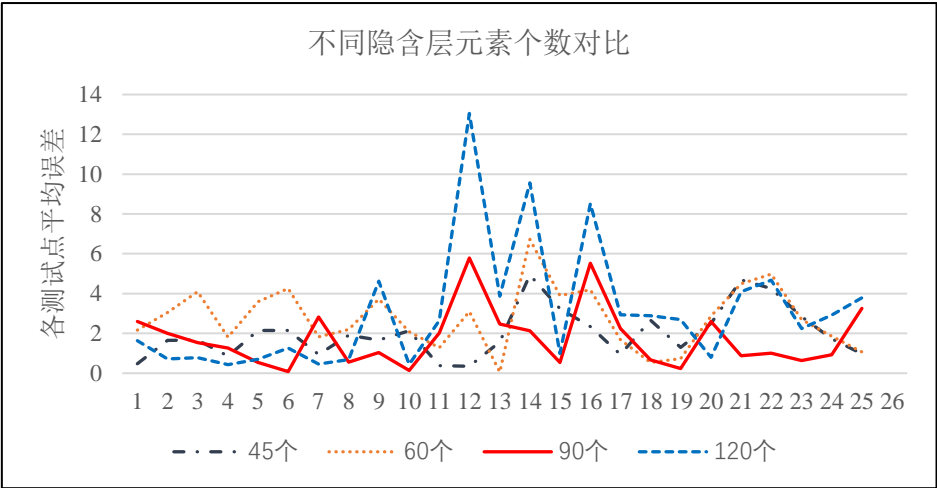


图 4 不同隐含层元素个数对应的各节点误差

从图 4 中的误差均值可以看出当隐含层元素个数为 90 时误差较为平稳且平均值最低。因此，设置隐含层元素个数为 90。在隐含层元素个数设定为 90 的条件下，分别设置学习率为：0.1,0.4,0.7,0.9。图 5 为不同学习率下的节点误差。结合表 3 中的误差均值可以看出，当学习率为 0.4 时误差最小。综上所述，该神经网络的迭代次数设置为 1000，训练目标设置为 0.000001，隐含层数设置为 90，学习率设置为 0.4。

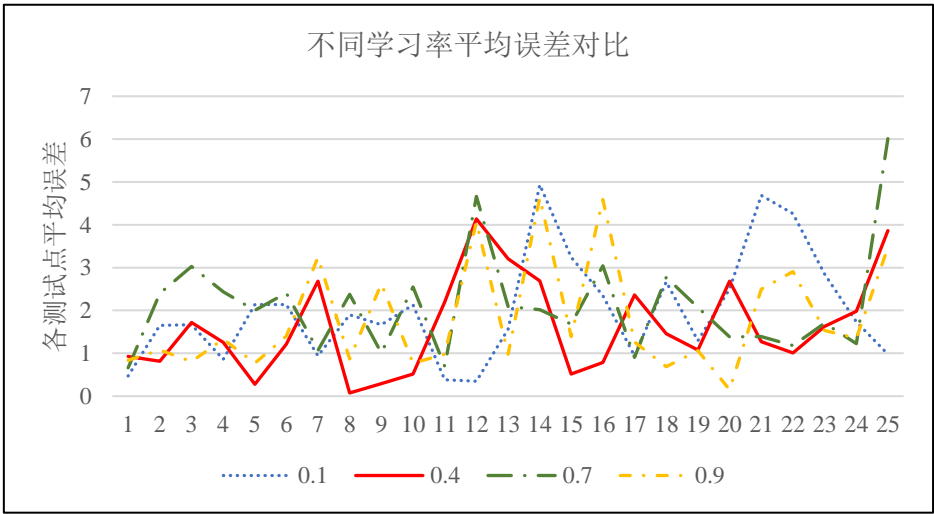


图 5 不同学习率误差对比

(3) 模型模拟

在上节设定的参数和训练完成的神经网络下，将 301-325 号的变量值带入网络得到模拟值。其中相关程度如图 6（a）所示，从数据中看出拟合度良好。拟合误差如图 6（b）所示，两个输出变量真实值和拟合值平均误差在 10 以内，预测精度拟合效果较好。模拟值和真实值曲线如图 7 所示。

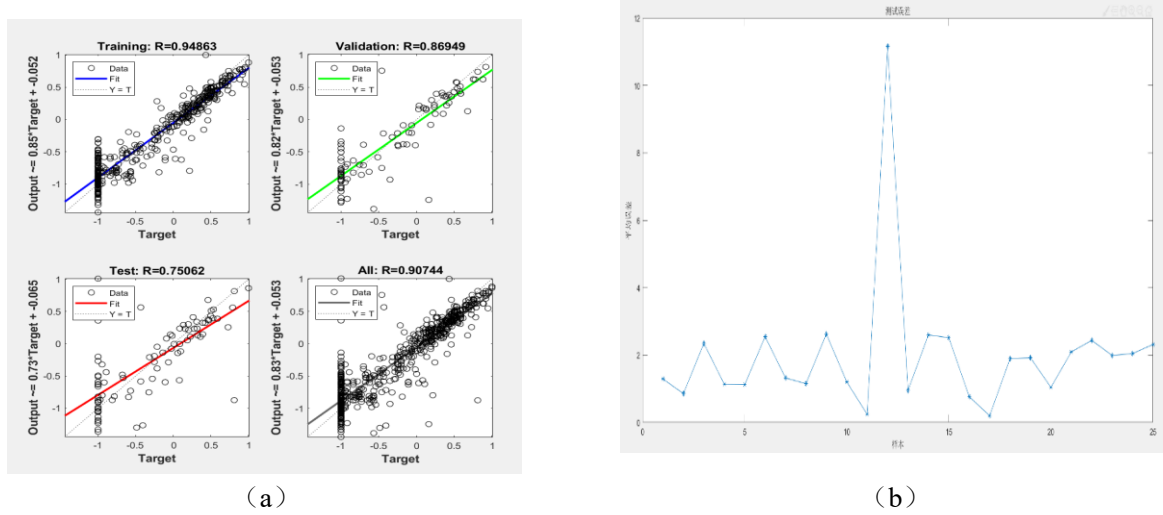


图 6 模型误差和拟合曲线

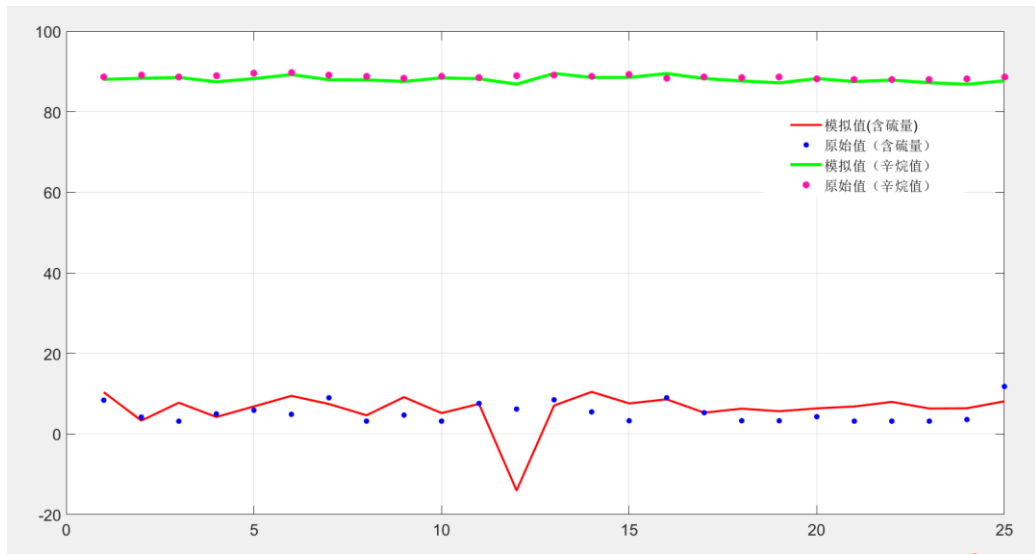


图 7 真实值和拟合值曲线

## 2.4 辛烷值损失优化模型构建

### 2.4.1 模型建立

根据优化目标，即在保证产品硫含量不大于  $5\mu\text{g/g}$  的前提下，辛烷值（RON）损失降幅大于 30%，且优化过程中原料、待生吸附剂、再生吸附剂的性质保持不变，本优化方案的目标函数及约束条件为：

$$f(x) = \min Z \quad (19)$$

$$s.t. \begin{cases} \frac{(Z - \bar{Z})}{\bar{Z}} < 30\% \\ 0 < S < 5 \\ (X_i)_t = (X_i)_{t-1}, X_i \text{ 为原料性质变量} \end{cases}$$

其中， $Z$  为辛烷值损失值， $\bar{Z}$ 为样本中辛烷值损失平均值， $S$  表示产品中硫含量。

#### 2.4.2 模型求解算法设计

本研究的问题属于优化问题，进而构建优化模型。遗传算法<sup>[15]</sup>在求解多变量优化问题具有较好的适用性，因此本研究拟通过遗传算法求解模型。算法总体流程如图 8 所示。

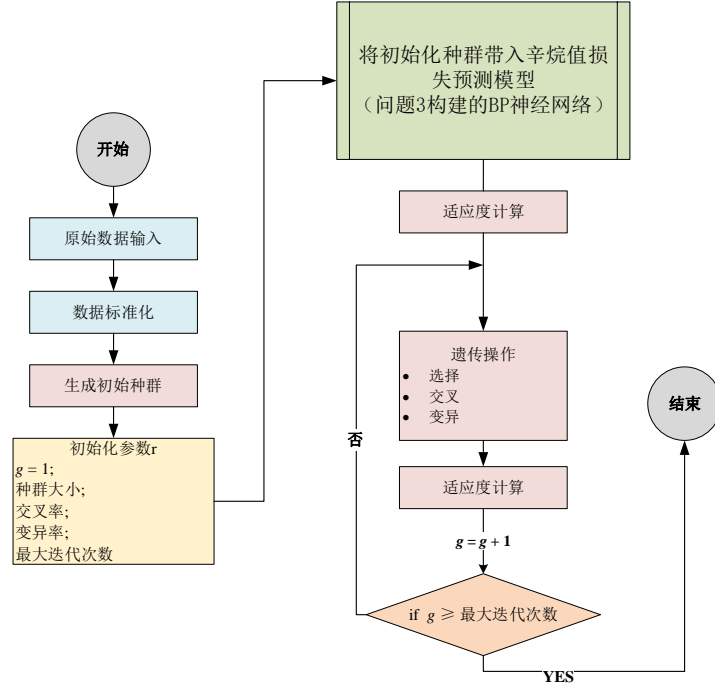


图 8 算法总体流程

本研究采用实数编码遗传算法对构建的数学模型进行计算。遗传算法是一种模拟自然进化的人工智能方法，通过算子选择、交叉和变异三个核心步骤计算产生适应度更高的个体。首先将整理好的原始数据进行标准化处理，按照设定的规则随机产生初始种群并设定遗传算法相关参数；将构建的初始种群带入构建的辛烷值损失预测模型（上节中构建的 BP 神经网络）；然后计算种群中所有个体的适应度，继而进入遗传操作进行种群更新并计算更新后种群所有个体的适应度；判断迭代次数是否达到设定的最大迭代次数和优化目标，如达到根据适应度函数选出种群中最优个体，输出各变量值。否则，对种群继续进行遗传操作。具体计算流程如下：

##### （1）数据归一化处理

归一化处理是消除各节点状态值量纲不同而引起的计算误差的过程。将数据归一化至 $[-1, 1]$ 范围内。本文采用离差标准化的形式对原始数据进行处理，见式（20）。其中 $x^*$ 是标准化后数据， $x$ 是原始数据。

$$x^* = \frac{x - \max}{\max - \min} \quad (20)$$

##### （2）染色体编码

由于实数编码更容易理解、节省基因操作时间且具有范围大和精度高的优点，因此本研究使用实数编码<sup>[16]</sup>的方式。本研究在第二问中筛选出 30 个变量，其中 2 个原材料性质变量（辛烷值、饱和烃），28 个操作变量。由于原材料性质变量不变，只能优化操作变量。染色体长度设置为 30，其中前两个染色体为性质变量，不进行交叉、变异等操作，为恒定值，后 28 个变量为操作变量，如图 9 所示。染色体



的个体值是在标准化后的取值范围内随机产生，因为本文标准化数据是将数值归一到[-1,1]区间内，因此除了染色体前两个个体根据实际值的标准化数据外，其他的个体均在（-1,1）范围内随机产生。

x1	0.2
x2	0.4
	...
x30	0.9

图 9 染色体编码示例

### (3) 初始化种群

构建过程中首先需要确定种群的大小为  $PZ$ ， $PZ$  个个体构成了一个种群，初始的种群为一个 30 行  $PZ$  列矩阵。

### (4) 适应度和选择操作

选择操作是从父代种群中根据计算的适应度值按照一定概率选择优良的个体组成新的种群。个体被选中的概率与适应度值有关，个体适应度越高被选中的概率越大。遗传算法的选择操作有精英选择、对抗法、锦标赛法和轮盘赌等多种方法，考虑计算效率本研究使用轮盘赌法进行选择操作。具体步骤如下：

**Step1:** 计算适应度  $err = R_0 - R_1$ 。首先将种群带入训练好的 BP 神经网络计算产成品辛烷值  $R_0$  和含硫量  $L_0$ 。由于本研究优化条件为辛烷值（RON）损失降幅小于 30%且越小越好，因此设定适应度函数为辛烷值（RON）损失值的倒数，见式（21）。其中  $R_0$  为原材料辛烷值， $R_1$  为产成品辛烷值， $\overline{R_{lost}}$  为辛烷值损失值均值。因为模型中约束条件为含硫量小于  $5 \mu g/g$  和辛烷值（RON）损失降幅大于 30%。所以设定不满足这两个条件，适应度函数赋值为 0.0001。其中  $R_1$  为优化后的产成品辛烷值。

$$fitness = \begin{cases} \frac{1}{R_0 - R_1} (L_0 < 5 \text{ 且 } (R_0 - R_1) / \overline{R_{lost}} < 130\%) \\ 0.0001 (L_0 > 5 \text{ 或 } (R_0 - R_1) / R_0 > 130\%) \end{cases} \quad (21)$$

**Step2:** 计算种群中个体被选择的概率。按式（22）计算种群中个体  $z$  被选中的概率  $P_z$ 。

$$P_z = \frac{fitness_z}{\sum_{z=1}^{PZ} fitness_z} \quad (22)$$

**Step3:** 计算个体  $z$  的累计概率  $Cp_z$ ，见式（23）：

$$Cp_z = \sum_{l=1}^z P_z \quad (23)$$

**Step4:** 确定被选中个体。在[0,1]区间内产生一个均匀分布的伪随机数  $r$ 。若  $r < Cp_k$ ，则种群中第  $k$

个个体被选中。否则选中第  $L$  个个体，满足  $Cp_{L-1} < r \leq Cp_L$ 。此后重复 step4，共重复  $PZ$  次。

#### (5) 交叉操作

交叉操作是从父代种群中随机选择两个个体，通过两个染色体交换组合，把父代的优良染色体遗传给子代进而产生新的优良个体。交叉方式有单点交叉、两点交叉、多点交叉和算数交叉，由于本研究采用实数编码且个体编码结构复杂，考虑到计算效率这里采用两点交叉的方式进行交叉操作。具体流程如下：

**Step1:** 在父代种群中随机选择两个个体作为父代个体，然后在(0,1)区间按照均匀分布随机产生一个随机数  $r$ ，如果  $r < P_{cross}$  则执行交叉操作（进入 Step2），否则原种群中被选中的两个父代个体不变化（重复执行 Step1），其中  $P_{cross}$  为交叉率（根据经验人工设定，取值范围为 (0,1)）。

**Step2:** 首先随机产生两个交叉点的位置， $pos1$  和  $pos2$ ，要求  $pos1 \neq pos2$ 。然后在选出的两个父代个体中交换  $pos1$  和  $pos2$  两点间的基因片段。如图 10 所示。此后重复 step1 和 step2，共重复  $PZ$  次。

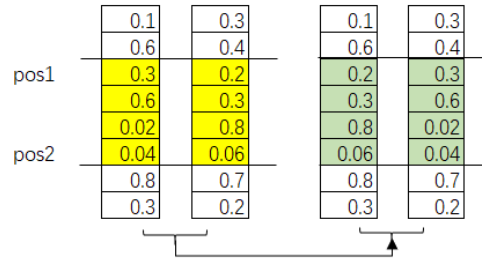


图 10 交叉操作示例

#### (6) 变异操作

变异操作的作用是维持种群的多样性，从种群中随机选取一个个体，选择个体中的一个或几个染色体进行变异以产生更优秀的个体。本研究使用针对实编码标准的多元非均匀变异算子进行变异操作<sup>[17]</sup>，其操作流程如下：

**Step1:** 在交叉变异后的种群中随机选择一个个体，产生一个均匀分布随机数  $r$ 。如果  $r < P_{mutation}$  该个体被选中执行变异操作（进入 Step2）；否则返回 Step1。

**Step2:** 在被选个体中，随机选择一个染色体进行变异操作（该个体是否变异取决于变异率，与交叉操作类似），见式 (24) (25)。表示种群中第  $j$  个个体第  $i$  行的元素。 $a_{max}$  和  $a_{min}$  分别表示基因的上下界， $g$  是当前迭代次数， $G_{max}$  是最大进化次数， $r_1, r_2$  为[0,1]区间随机数。此后重复 step1 和 step2，共重复  $PZ$  次。

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{max}) \cdot f(g), & r_1 \geq 0.5 \\ a_{ij} + (a_{min} - a_{ij}) \cdot f(g), & r_1 < 0.5 \end{cases} \quad (24)$$

$$f(g) = r_2 \cdot \left( 1 - \frac{g}{G_{max}} \right)^2 \quad (25)$$

#### 2.4.3 遗传参数确定

本研究采用的遗传算法需要设定相应的遗传参数，主要包括种群规模、最大迭代次数、变异率和交叉率。

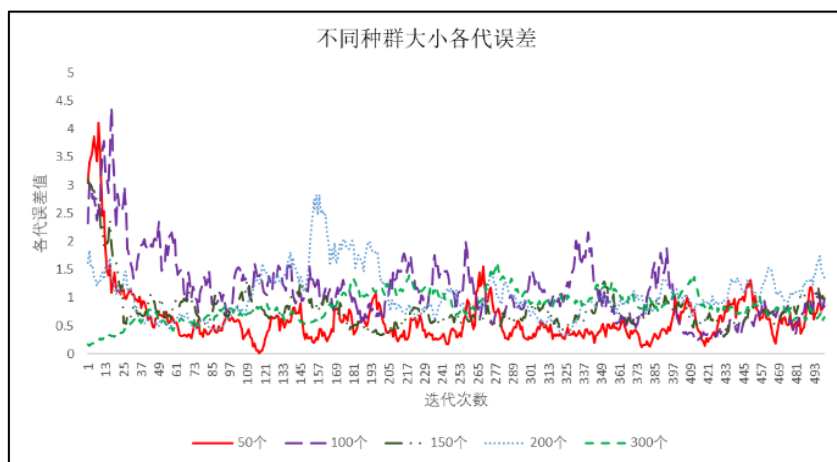
确定种群规模：在最初实验中，种群规模分别设置为 50, 100, 150, 200, 300。其他遗传参数交叉率设置为 0.5, 变异率设置为 0.05, 最大迭代次数设置为 500。按照设定的种群规模分别进行 5 次运算，对比 5 次运算各代最优个体均方差的平均值。从图 11 (a) 中的计算结果不难看出，当种群规模为 50 时个体的误差最小，在降幅在 30% 条件下的误差计算公式见式 (26)，因此本研究种群规模设置为 50。

$$err = R_0 - R_i \quad (26)$$

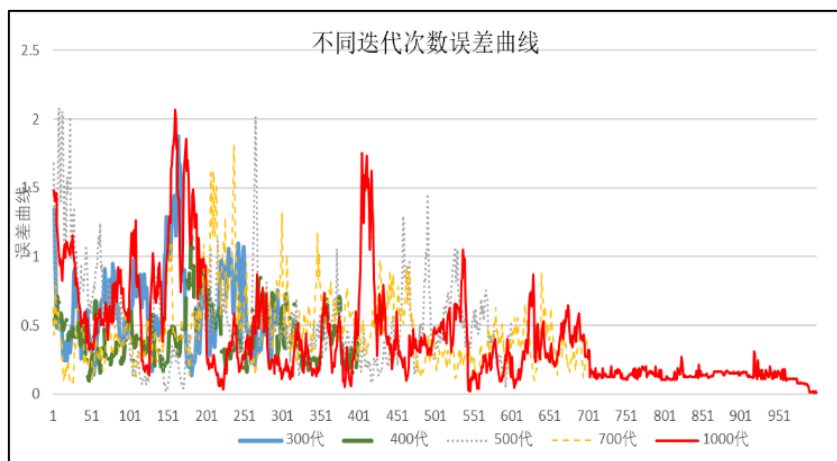
确定最大迭代次数：最大迭代次数分别设置为 300, 400, 500, 700, 1000。种群规模设置为 50, 交叉率设置为 0.5, 变异率设置为 0.05。同样按照设置的各迭代次数分别计算 5 次，通过各代最优个体均方差的均值对最大迭代次数参数进行讨论。当最大迭代次数为 1000 代时误差最小，如图 11 (b) 所示。因此设置最大迭代次数为 1000 代。

确定交叉率：交叉率分别设置为 0.1, 0.2, 0.5, 0.6, 0.7, 种群规模设置为 50, 最大迭代次数设置为 1000, 变异率设置为 0.05。按照设置的交叉率各计算 5 次分别进行讨论。当交叉率为 0.2 时是所有备选方案中最优的，见图 11 (c)，所以交叉率设置为 0.2。

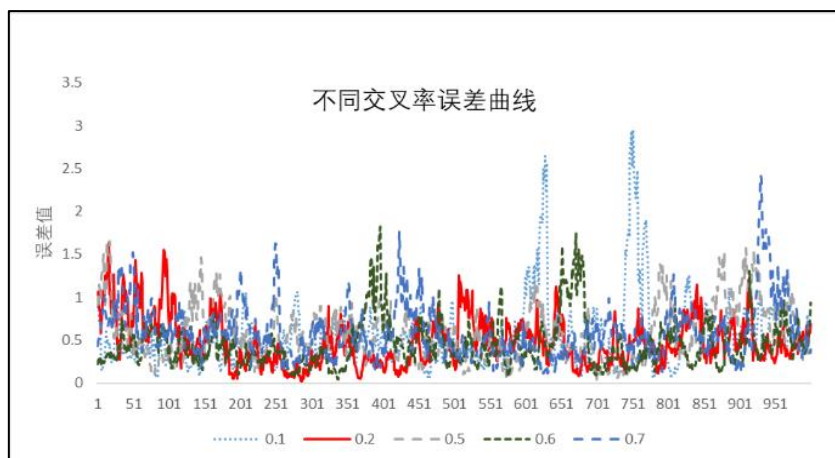
确定变异率：以避免陷入局部最优解，变异率其值一般都在 0.1 以下，因此变异率备选方案设为 0.01, 0.02, 0.05, 0.06, 0.09, 迭代次数设置为 1000, 种群规模设置为 50, 交叉率设置为 0.2。按照上述同样的方法对变异率进行讨论，当变异率设置为 0.09 时达到最优，见图 11 (d)。



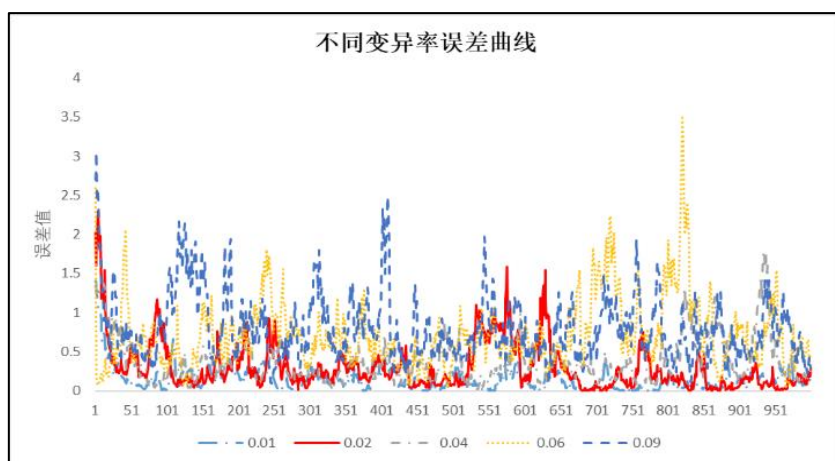
(11-a)



(11-b)



(11-c)



(11-d)

图 11 遗传参数设定

#### 2.4.4 优化操作

通过对附件一中的辛烷值损失量的变化幅度测算，超过 30%幅度的数据共有 30 组数据。其数据编号和相关变量值见表 4。基于构建的优化模型，分别将各组数据的原材料的辛烷值和饱和烃数据作为构建的优化模型染色体的前两个变量，其他变量在其相应的区间内随机产生，进行优化操作，进而计算得到各操作变量的优化值，以满足产成品的硫含量小于  $5 \mu\text{g/g}$ ，辛烷值损失量最少。优化后的产成品的辛烷值、硫含量、辛烷值损失量和各操作变量值见表 5 所示。从表中可以看出经过优化后的产成品辛烷值大幅度缩减且含硫量均小于  $5 \mu\text{g/g}$ 。从辛烷值损失值优化缩减非常大，实现了非常好的效果。这属于理论上的理想情况，但是在实际应生产中还存在误差。通过对优化操作变量表 5 中的数据分析，优化后的各操作变量指标的取值范围见表 6，即若想实现产成品在含硫量小于  $5 \mu\text{g/g}$  的限制条件下，产成品辛烷值达到最优，应保证这些节点的数据在该范围内。

表 4 辛烷值降幅超过 30%的各样本及部分变量值

样本 编号	时间	原料性质		产品性质		
		辛烷值 RON	饱和烃, % (烷 烃+环烷烃)	硫含量, µg/g	辛烷值 RON	RON 损失
12	2020/4/21 8:00:00	90.2	53.00	3.2	88.90	1.3
13	2020/4/16 8:00:00	89.2	55.50	6.3	87.70	1.5
23	2020/3/13 8:00:00	89.6	56.59	3.2	87.99	1.61
59	2019/11/5 8:00:00	89.4	63.40	3.2	87.65	1.75
61	2019/10/29 8:00:00	89.7	61.60	3.2	88.02	1.68
62	2019/10/25 8:00:00	89	59.94	5.8	87.22	1.78
63	2019/10/22 8:00:00	89.1	58.70	3.2	87.52	1.58
64	2019/10/18 8:00:00	88.8	58.81	3.2	87.22	1.58
69	2019/9/30 8:00:00	88	58.50	3.2	86.42	1.58
70	2019/9/27 8:00:00	88.7	57.30	3.2	86.92	1.78
71	2019/9/24 8:00:00	90.6	56.50	3.2	88.82	1.78
76	2019/9/3 8:00:00	88.7	62.20	3.2	86.9	1.8
79	2019/8/23 8:00:00	89.9	53.40	4.7	88.22	1.68
82	2019/8/13 13:52:40	88.7	56.80	3.2	87.12	1.58
83	2019/8/9 8:00:00	89.3	56.40	3.2	87.72	1.58
84	2019/8/6 8:00:00	87.6	56.20	3.2	85.82	1.78
91	2019/7/6 10:25:06	90.1	53.25	3.2	88.52	1.58
98	2019/6/18 8:00:00	91.3	48.40	3.2	90.1	1.2
107	2019/5/14 8:00:00	91.2	49.30	3.2	89.52	1.68
108	2019/5/10 8:00:00	90.6	50.40	4.5	88.92	1.68
114	2019/4/5 8:00:00	91.3	51.60	3.7	89.72	1.58
121	2019/3/12 8:00:00	90.6	54.30	3.2	89	1.6
131	2019/2/12 8:00:00	89.4	55.90	3.2	87.8	1.6
135	2019/2/4 8:00:00	89.2	59.30	3.2	87.39	1.81
138	2019/1/28 8:00:00	89.3	59.30	3.2	87.59	1.71
149	2019/1/2 8:00:00	88.2	56.90	3.2	86.4	1.8
151	2018/12/28 8:00:00	87.8	56.30	3.2	85.99	1.81
153	2018/12/24 8:00:00	87.5	56.30	3.2	85.68	1.82
173	2018/7/09 8:00:00	89.8	56.10	3.2	88.08	1.72
303	2017/6/6 12:36:04	90.3	43.24	3.2	88.65	1.65

表 5 优化后的各操作变量值、产品辛烷值、含硫量和辛烷值损失

样本编号	产成品			原材料		操作变量																											
	硫含量	辛烷值	RON损失值	辛烷值RON	饱和烃,wt% (烷烃+环烷烃)	还原器压力	稳定塔顶压力	塔气出装量温度	蒸汽进装置压力	10MPa蒸汽进装置温度	非净化风进装置压力	净化风进装置压力	3#催化汽油进装置流量	混氢点氢气流量	加热炉排烟温度	反应器入口温度	D104压力	再生器温度	E-101D壳程出口管温度	E-101C管程出口管温度	D-124液位	D-122液位	再生器下部温度	再生器下部温度	D-113压力	低压热氮气压力	R-102床层吸附剂料位密度	R-102 #1通风挡板温度	K-102A进口温度	F-101出口总管压力	F-101出口支管#4温度	F-101出口支管#3温度	F-101出口支管#1温度
12	4.92	90.20	0.0010	90.60	53.23	2.5174	0.6555	26.8753	0.9983	210.3687	0.6048	0.4614	84.7187	6400.5780	158.3837	413.8832	2.2882	489.9026	127.8857	360.3140	0.2140	-1.2630	509.0944	468.7028	0.0657	0.2911	77.7752	469.6581	25.3339	2.4801	416.7349	416.3484	414.5950
13	0.74	89.20	0.0001	89.20	55.50	2.5519	0.6421	30.5786	0.8768	186.7848	0.6096	0.5206	66.1913	7309.4094	121.7070	409.4569	2.4158	506.3048	89.3851	375.6413	6.7521	-1.2526	467.0024	494.7052	0.0841	0.3672	54.2353	481.5403	23.7188	2.4504	414.3836	408.4991	407.1165
23	0.15	89.20	0.4002	89.60	56.59	2.4541	0.6600	34.9692	1.0658	228.7772	0.6453	0.4823	59.0005	7658.6170	82.2853	416.1676	2.3365	507.2820	105.7717	380.4727	0.6505	-1.2535	492.1634	506.0180	0.0992	0.4040	79.6796	502.9559	22.8839	2.4471	405.3374	416.1487	411.6971
59	2.46	89.40	0.0001	89.40	63.40	2.4027	0.6462	37.5585	0.7072	192.3998	0.6553	0.5285	50.0608	8477.0709	136.5132	413.6070	2.3342	493.9123	110.0099	383.4625	-1.5059	-1.2618	504.6964	487.1543	0.0811	0.3969	85.3364	503.7970	22.3255	2.5067	415.7294	410.2634	411.9959
61	2.11	89.70	0.0001	89.70	61.60	2.4657	0.6515	30.8797	0.5755	229.9326	0.5625	0.4568	38.4244	7571.1766	27.6446	403.4613	2.3209	474.4871	127.6731	363.8579	-1.5908	-1.2639	470.6333	472.5177	0.0920	0.3436	87.1647	484.4338	25.6922	2.4310	417.3809	411.4121	411.0162
62	0.53	89.00	0.0011	89.00	59.94	2.4244	0.6431	27.3128	0.7777	185.4728	0.5917	0.4134	79.2859	6752.5208	74.4829	416.5399	2.2902	472.4586	109.3014	374.8964	6.7059	-1.2524	469.2348	495.5945	0.0866	0.3001	89.1793	479.4539	24.8073	2.3866	408.7487	420.4946	412.8396
63	2.94	89.10	0.0005	89.10	58.70	2.4116	0.6521	28.1611	0.6925	222.4854	0.6281	0.4705	17.9028	7662.8225	28.3555	413.2483	2.2364	502.5238	94.9109	386.6624	4.8088	-1.2613	498.4428	496.3626	0.0554	0.3466	56.8203	498.2531	24.1823	2.5007	410.8497	413.7827	407.2286
64	4.76	88.80	0.0002	88.80	58.81	2.5486	0.6550	25.2123	1.1043	236.3910	0.6188	0.4079	71.9544	7083.9723	43.0366	408.1006	2.2784	511.2604	92.6705	368.7770	0.8237	-1.2544	489.4936	468.8041	0.0743	0.3356	82.0636	495.0928	23.2568	2.5992	406.5529	418.4198	410.6773
69	3.06	88.00	0.8004	88.80	58.81	2.4954	0.6416	23.8055	0.8386	179.2344	0.5713	0.4391	63.5923	8622.8409	31.0834	408.4896	2.2965	509.0032	97.4960	362.3085	0.9930	-1.2522	475.2975	495.2592	0.0785	0.3717	61.8920	484.3503	24.3425	2.4445	410.6049	416.6129	410.1156
70	1.90	88.70	0.0001	88.70	57.30	2.5862	0.6547	28.8697	0.9854	240.9034	0.5719	0.4242	38.6711	6326.0145	75.2787	410.8779	2.3842	476.6763	91.8387	378.0846	0.3855	-1.2495	468.2819	484.3944	0.0540	0.3171	55.4498	473.7991	22.6765	2.4764	414.0260	416.4814	413.7260
71	2.96	90.60	0.0001	90.60	56.50	2.4821	0.6475	24.7428	1.0729	178.4980	0.5780	0.4451	51.4541	8751.7939	35.6623	404.7068	2.2222	471.4608	96.5931	363.7495	-1.1899	-1.2549	498.8418	502.5389	0.0658	0.4086	61.9365	494.6989	24.2160	2.4484	405.7540	420.4209	406.4901
76	1.59	88.70	0.0001	88.70	62.20	2.4458	0.6573	29.9637	0.7595	224.6252	0.5968	0.4903	4.9505	7543.0129	132.7197	405.6049	2.3181	487.7914	121.9079	388.7062	1.7168	-1.2590	509.6049	466.7427	0.0963	0.2811	70.9983	482.6488	21.9463	2.3854	407.0570	411.6734	409.6514
79	4.11	89.90	0.0005	89.90	53.40	2.3890	0.6547	32.3963	0.6621	182.0928	0.6622	0.4392	2.8526	5934.1445	131.2494	414.5338	2.3392	509.9557	112.2151	377.4175	0.7415	-1.2637	512.3077	506.9276	0.0643	0.3378	83.7715	502.8507	25.2669	2.4875	408.5662	417.2554	409.0279
82	3.43	88.70	0.0002	88.70	56.80	2.5303	0.6414	31.9688	0.5799	183.2881	0.6129	0.5492	43.8606	7721.5141	111.9266	406.2800	2.3095	484.6511	127.3519	375.3477	0.2173	-1.2598	484.4527	473.9990	0.0546	0.2666	48.9289	488.0713	24.4697	2.4545	405.7754	414.7098	410.2921
83	2.36	89.30	0.0002	89.30	56.40	2.4725	0.6569	24.5354	0.5818	184.9810	0.6418	0.4888	6.5995	5983.0659	101.1310	415.3061	2.3468	484.6727	123.6462	362.7341	2.2026	-1.2511	469.8801	508.5946	0.0603	0.3169	47.6138	472.3223	22.7991	2.5194	405.0760	413.0022	411.0493
84	4.14	87.60	0.0002	87.60	56.20	2.5576	0.6544	37.6525	1.0994	214.3184	0.5801	0.4154	68.7221	6510.3877	127.7022	405.9157	2.2465	485.7048	82.7253	356.3627	5.2791	-1.2519	490.7558	469.4858	0.0516	0.3697	72.0055	480.7680	22.1598	2.4700	411.3127	409.7769	414.5687
91	2.77	90.10	0.0001	90.10	53.25	2.5019	0.6510	30.5249	0.9473	247.7871	0.6625	0.5641	83.2047	7900.0369	144.3742	404.3391	2.2539	512.3060	88.8486	360.6564	4.0212	-1.2584	509.0066	506.0903	0.0507	0.2947	56.7007	509.1459	25.8240	2.4876	410.5748	418.6235	411.2189
98	3.24	91.30	0.0000	91.30	48.40	2.5762	0.6502	36.6843	1.2311	217.6645	0.5550	0.5534	29.3131	6731.2208	106.6762	416.2343	2.3477	492.7960	105.3507	369.8099	0.8381	-1.2543	509.5223	507.8660	0.0980	0.2644	67.2417	500.5439	24.5902	2.4204	410.0742	418.0558	408.9460
107	2.52	91.20	0.0002	91.20	49.30	2.5783	0.6467	33.3945	0.9884	190.1021	0.6591	0.5015	37.9870	8770.5981	152.5008	403.5925	2.3053	485.6615	126.6889	376.9657	5.2835	-1.2535	502.2146	472.2903	0.0935	0.2859	85.0822	494.9141	26.1385	2.5461	416.7608	419.4997	412.2994
108	4.08	90.60	0.0001	90.60	50.40	2.5329	0.6552	23.9861	1.1429	203.6349	0.6526	0.4063	62.8185	7077.7527	126.3739	411.9368	2.4038	496.8512	99.9470	367.6894	1.1419	-1.2535	508.7055	501.5908	0.0694	0.3118	81.9270	494.4627	23.6990	2.4916	405.4109	416.0456	412.7484
114	4.27	91.30	0.0006	91.30	51.60	2.3929	0.6488	37.5735	1.2627	211.1485	0.6518	0.4508	75.2918	8859.4644	107.9739	414.4820	2.2182	483.1471	85.4321	372.4040	6.6766	-1.2626	492.3081	500.0282	0.0654	0.2912	84.5842	510.0917	25.8712	2.4768	404.4955	413.3288	415.1301
121	1.76	90.60	0.0004	90.60	54.30	2.4804	0.6546	34.5235	0.5863	211.7083	0.6435	0.4981	80.7627	8050.7709	137.3408	407.0272	2.2718	514.3529	122.7665	376.4543	4.7349	-1.2507	472.1191	478.5899	0.0581	0.3953	80.1810	498.5559	25.0331	2.4287	411.7430	420.4821	409.3297
131	3.31	89.40	0.0003	89.40	55.90	2.4069	0.6459	32.7308	0.7901	205.0036	0.6246	0.4919	9.3202	8829.9097	102.5093	404.1864	2.2620	487.4247	99.1223	375.7455	1.9835	-1.2548	487.7259	471.7590	0.0579	0.3611	52.9565	473.7718	23.6202	2.3913	410.6636	414.3278	417.7095
135	3.23	89.20	0.0001	89.20	59.30	2.5375	0.6454	28.9326	0.6745	205.7838	0.6225	0.4887	43.4501	8603.5842	132.9777	408.7949	2.4331	482.3743	123.2470	388.1245	5.7519	-1.2597	507.6819	494.6568	0.0926	0.3580	56.6696	462.3794	25.3426	2.5553	407.5014	419.4820	416.8809
138	0.17	89.30	0.0000	89.30	59.30	2.5162	0.6467	29.9556	1.2127	241.2753	0.6271	0.5615	4.6600	7046.8259	44.0068	407.0824	2.2597	500.8950	101.8269	358.7887	3.9774	-1.2575	504.5775	491.5019	0.0517	0.3109	70.8926	475.7658	25.1911	2.4866	404.3582	414.5470	406.4404
149	3.76	88.20	0.0000	88.20	56.90	2.4849	0.6527	32.8919	0.7277	246.7475	0.6756	0.4085	7.2079	6088.2308	257.7251	410.4804	2.3661	494.0148	104.2831	372.4100	-0.1131	-1.2490	499.6161	465.5800	0.0993	0.2809	54.2090	476.5598	25.0191	2.3992	408.1670	416.4072	409.9875
151	1.48	87.80	0.0004	87.80	56.30	2.4270	0.6537	33.5764	1.1226	186.6313	0.6664	0.4327	17.8531	8288.2388	29.2741	415.0487	2.2471	485.6679	97.8611	383.1736	4.6217	-1.2551	484.6615	499.7301	0.0645	0.3090	54.8570	486.8107	25.5937	2.4664	410.1115	412.4928	412.2137
153	1.40	87.50	0.0001	87.50	56.30	2.5138	0.6404	32.9102	1.1751	213.1527	0.5588	0.4336	85.0803	8835.0515	26.2852	412.2892	2.3531	488.7577	124.5234	356.6104	4.5720	-1.2552	474.8953	500.6877	0.0949	0.3457	72.8067	496.5232	23.9602	2.3945	411.3271	414.4261	416.0099
173	4.82	89.80	0.0010	89.80	56.80	2.4291	0.6471	31.8517	1.2138	178.3804	0.6665	0.5628	5.2974	8091.2710	65.8888	410.5962	2.3980	495.2367	99.8255	377.6998	-0.3539	-1.2498	478.9795	502.8232	0.0950	0.4076	67.4003	501.7843	23.0813	2.4581	414.7154	418.0782	405.4355
303	2.80	90.30	0.0003	90.30	43.24	2.5658	0.6561	37.0407	0.5285	212.9314	0.5697	0.5505	84.8838	8341.4375	68.7125	412.7888	2.2835	491.3685	111.4796	365.2720	2.4316	-1.2617	474.1741	479.4012	0.0521	0.3104	53.9791	466.1517	23.6245	2.4477	411.1287	414.4879	412.

表 6 优化后各操作变量指标范围

操作变量指标	还原器 压力	稳定塔 顶压力	干气出装 置温度	蒸汽进装 置压力	1.0MPa 蒸汽进装 置温度	非净化风进 装置压力	净化 风进 装置 压力	3#催化汽 油进装置 流量	混氢点氢 气流量	加热炉排烟 温度
最小值	2.38899 91	0.64041 04	23.805546	0.5284571	178.3804 4	0.5550404	0.4063 242	2.8525845	5934.1445	26.285158
最大值	2.58617 06	0.65995 7	37.652489	1.2627181	247.7871 2	0.6755903	0.5641 444	85.080342	8859.4644	158.38368
操作变量指标	反应器 入口温 度	D104 压 力	再生器温 度	E-101D 壳 程出口管 温度	E-101C 管程出口 管温度	D-124 液位	D-122 液位 液位	再生器下 部温度	再生器下 部温度	D-113 压力
最小值	403.461 32	2.21824 71	471.46085	82.725276	356.3627	-1.590839	- 1.2639 05	467.00243	465.57997	0.0507429
最大值	416.539 88	2.43305 84	514.35287	127.88569	388.7062 1	6.7520683	- 1.2490 45	512.30769	508.59461	0.0993448
操作变量指标	低压热 氮气压 力	R-102 床层吸 附剂料 位密度	R-102 #1 通风挡板 温度	K-102A 进 气温度	F-101 出 口总管压 力	F-101 出口 支管#4 温 度	F-101 出口 支管 #3 温 度	F-101 出 口支管#1 温度		
最小值	0.26443 38	47.6138 42	462.37939	21.946257	2.385383 7	404.35815	408.49 914	405.4355		
最大值	0.40864 1	89.1792 68	510.09168	26.138528	2.599189 7	417.38089	420.49 463	417.70946		

## 2.5 模型可视化展示

### 2.5.1 数据指标处理

基于问题二中选出的关键变量指标，操作变量 28 个，原材料属性变量 2 个。其中原材料辛烷值和饱和烃值为原材料属性变量，因此不做调整。本研究以原材料辛烷值 90.06、饱和烃 53.23 为原材料性质变量值。通过调整属性变量的增加值测试产成品硫含量和辛烷值的变化，如表 7 所示。由于提供给出的数据中变量编号 139 样本的  $\Delta$  值为空值，本文假设为 0.1。



表 7 变量指标设置及变化区间

变量编号	中文名称	取值范围	Δ 值
0	原材料辛烷值	原材料真实值	
0	饱和烃,v% (烷烃+环烷烃)	原材料真实值	
3	还原器压力	2.35-2.70	0.1
15	稳定塔顶压力	0.60-0.70	0.05
19	干气出装置温度	20-40	1
23	蒸汽进装置压力	0.5-1.3	0.1
34	1.0MPa 蒸汽进装置温度	150-250	1
40	非净化风进装置压力	0.55-0.70	0.05
42	净化风进装置压力	0.35-0.60	0.05
48	3#催化汽油进装置流量	0-90	5
53	混氢点氢气流量	5500-9000	100
58	加热炉排烟出口温度	10-180	1
62	反应器入口温度	400-450	1
67	D104 压力	2.20-2.50	0.1
83	再生器温度	450-520	1
120	E-101D 壳程出口管温度	50-150	1
121	E-101C 管程出口管温度	300-400	1
132	D-124 液位	-1.8-(7.0)	0.5
139	D-122 液位液位	-1.5-(-1.2)	0.1
180	再生器下部温度	400-550	10
181	再生器下部温度	450-550	1
194	D-113 压力	0-0.15	0.05
215	低压热氮气压力	0.25-0.45	0.1
225	R-102 床层吸附剂料位密度	40-100	10
231	R-102 #3 通风挡板温度	300-500	5
273	K-102A 进气温度	20-30	1
289	F-101 出口总管压力	2.35-2.70	0.1
290	F-101 出口支管#4 温度	400-420	1
291	F-101 出口支管#3 温度	400-450	1
293	F-101 出口支管#1 温度	400-420	1

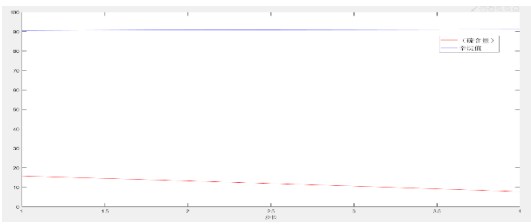
### 2.5.2 数据预测仿真可视化

将表 7 中的指标数据代入问题三中构建的辛烷值 BP 神经网络预测模型，从各变量的最小值开始，逐步按照标准化后的 Δ 值为步长进行调整。

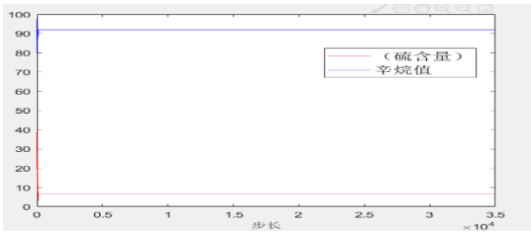
以操作变量“还原器压力”为例，将其他操作变量设定为最小值，令“还原器压力”从 2.35 值开始，以 Δ 值为步长，直到增加到最大值。将上述变量作为构建的 BP 神经网络预测模型的输入层变化量进行标准化，带入训练好的神经网络，得到产成品的辛烷值和硫含量值的标准化值，然后进行反标准化处理得到真实值。最后将调整过程中产成品的硫含量和辛烷值含量的变化过程进行可视化展现。图 12（1-28）分别为不同操作变量变化对产品硫含量和辛烷值影响的曲线。表 8 列示了各主要操作变量变化对辛烷值和硫含量的影响，主要分为正



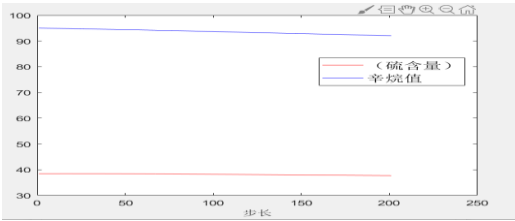
向影响、负向影响和正负变化较稳定的影响。



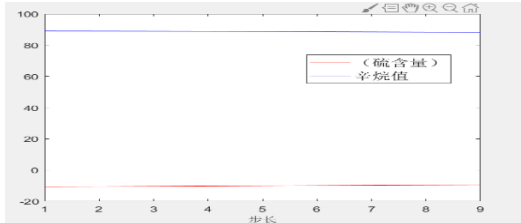
12-1（还原器压力变化）



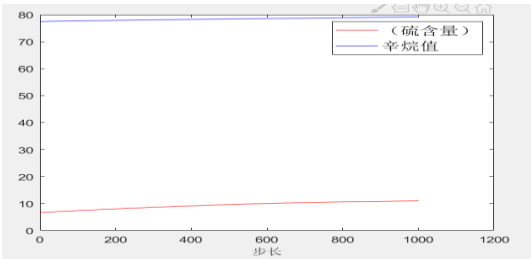
12-2（稳定塔顶压力变化）



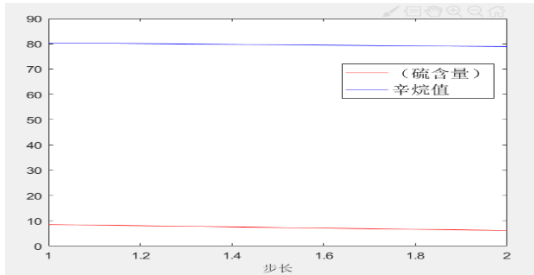
12-3（干气出装温度变化）



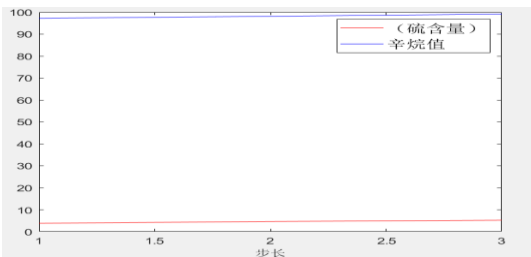
12-4（蒸汽进装置压力变化）



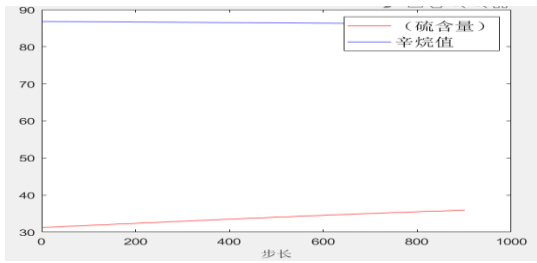
12-5（1.0MPa 蒸汽进装置温度变化）



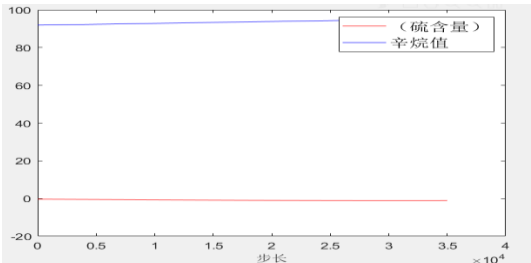
12-6（非净化风进装置压力变化）



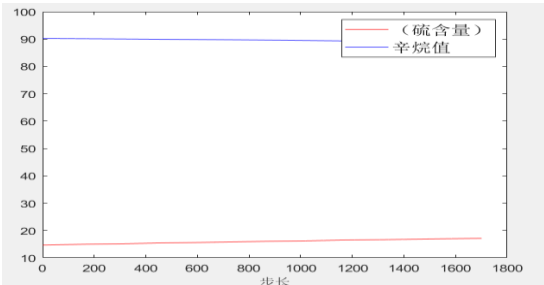
12-7（净化风进装置压力变化）



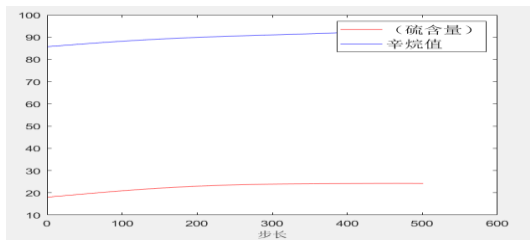
12-8（3#催化汽油进装置流量变化）



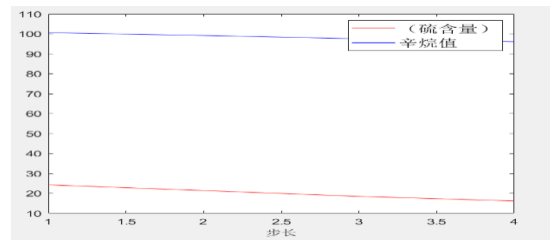
12-9（混氢点氢气流量变化）



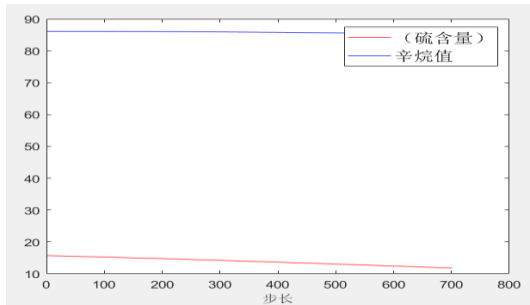
12-10（加热炉排烟出口温度变化）



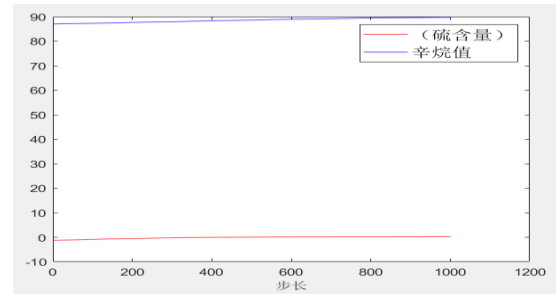
12-11 (反应器入口温度变化)



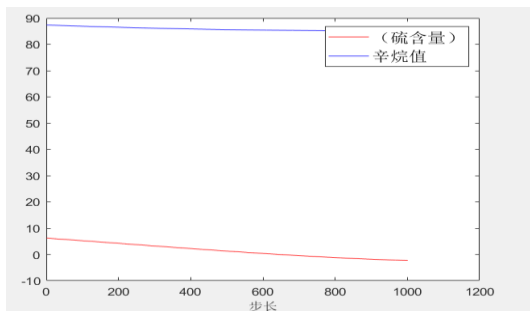
12-12 (D104 压力变化)



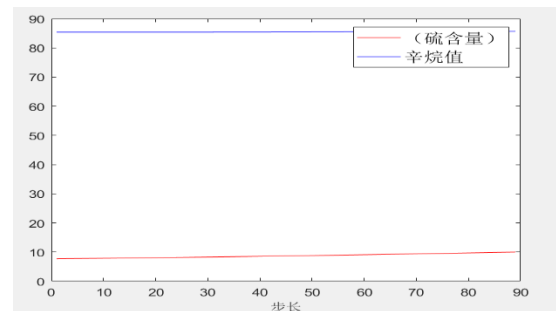
12-13 (再生器温度变化)



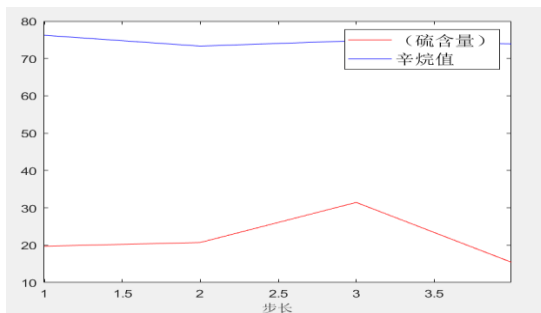
12-14 (E-101D 壳程出口管温度变化)



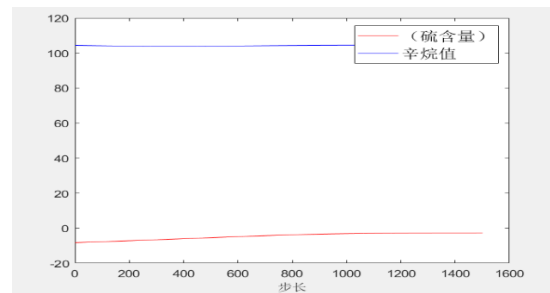
12-15 (E-101C 管程出口管温度变化)



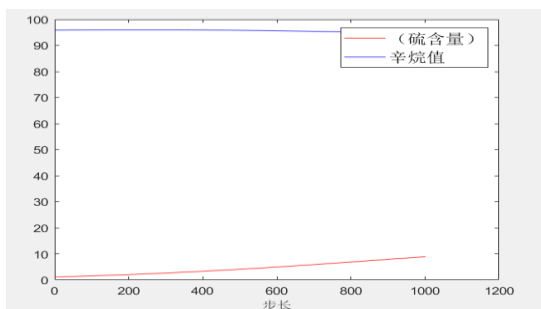
12-16 (D-124 液位变化)



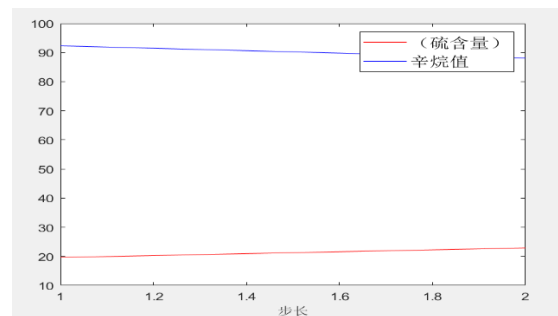
12-17 (D-122 液位变化)



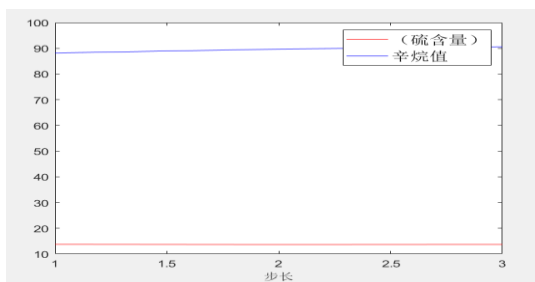
12-18 (再生器下部温度变化)



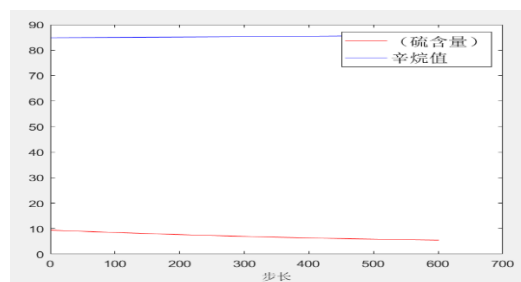
12-19 (再生器下部温度变化)



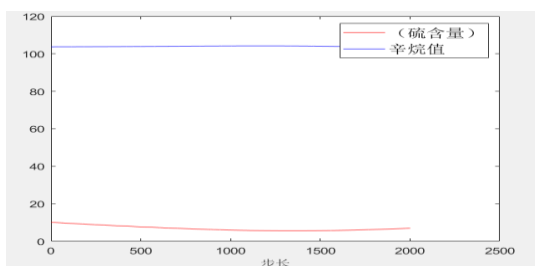
12-20 (D-113 压力变化)



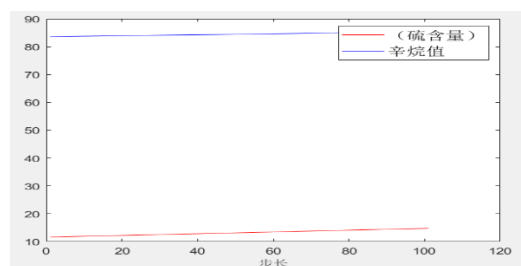
12-21(低压热氮气压力变化)



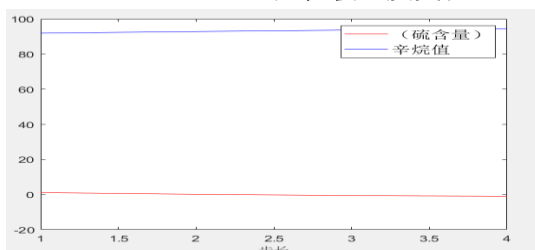
12-22(R-102 床层吸附剂料位密度变化)



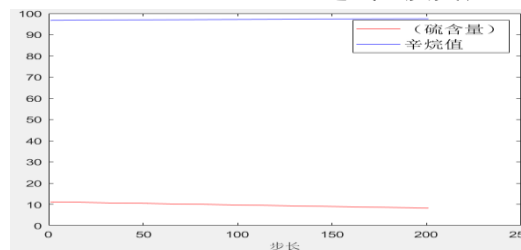
12-23 (R-102 #3 通风挡板温度变化)



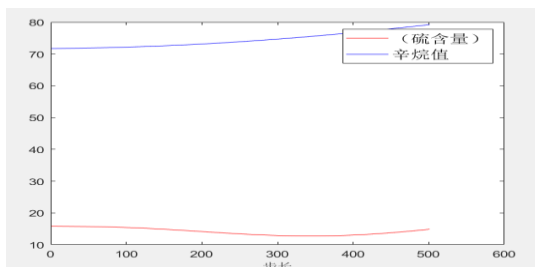
12-24 (K-102A 进气温度变化)



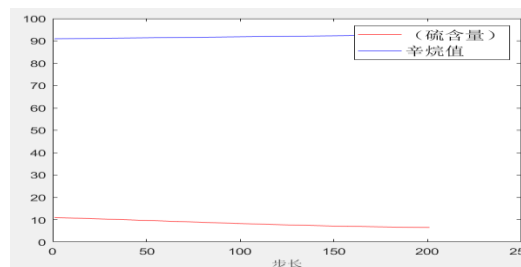
12-25 (F-101 出口总管压力变化)



12-26 (F-101 出口支管#4 温度变化)



12-27 (F-101 出口支管#3 温度变化)



12-28 (F-101 出口支管#1 温度变化)

图 12 不同操作变量变化对产成品辛烷值和含硫量的影响可视化图

表 8 各主要操作变量变化对辛烷值和硫含量的影响

	辛烷值			硫含量		
	正向影响	负向影响	稳定影响	正向影响	负向影响	稳定影响
还原器压力			√		√	
稳定塔顶压力			√			√
干气出装置温度						√
蒸汽进装置压力			√			√
1.0MPa 蒸汽进装置温度	√			√		
非净化风进装置压力			√		√	
净化风进装置压力	√			√		
3#催化汽油进装置流量		√		√		
混氢点氢气流量	√					√
加热炉排烟出口温度		√		√		
反应器入口温度	√			√		
D104 压力		√		√		
再生器温度			√		√	
E-101D 壳程出口管温度	√			√		
E-101C 管程出口管温度		√			√	
D-124 液位			√	√		
D-122 液位液位	√			√		
再生器下部温度	√			√		
再生器下部温度		√		√		
D-113 压力		√		√		
低压热氮气压力	√				√	
R-102 床层吸附剂料位	√				√	
密度						
R-102 #3 通风挡板温度			√	√		
K-102A 进气温度	√			√		
F-101 出口总管压力	√					√
F-101 出口支管#4 温度			√		√	
F-101 出口支管#3 温度	√					√
F-101 出口支管#1 温度	√					√

从图 12-1 可以看出，随着还原器压力的增加，产成品的含硫量呈现下降趋势，且产成品的辛烷值变化量不大；

从图 12-2 可以看出，还稳定塔顶压力变化对产成品的含硫量和辛烷值变化影响均不大；

从图 12-3 可以看出，干气出装置温度增加对产成品的辛烷值表现出负向的影响，对产成品的硫含量影响不大；

从图 12-4 可以看出，蒸汽进装置压力对产成品的含硫量和辛烷值的变化影响均不大；

从图 12-5 可以看出，1.0MPa 蒸汽进装置温度的增加使得产成品的辛烷值和硫含量均增加；

从图 12-6 可以看出，非净化风进装置压力的增加使得产成品的硫含量减少，但对辛烷

值变化不大，相对较为平稳；

从图 12-7 可以看出，净化风进装置压力增加使得产成品的辛烷值和硫含量均增加；

从图 12-8 可以看出，3#催化汽油进装置流量增加使得产成品硫含量增加，对辛烷值影响较小，呈微弱的负向影响；

从图 12-9 可以看出，混氢点氢气流量的增加提升了产成品的辛烷值量，但对硫含量的影响不大；

从图 12-10 可以看出，加热炉排烟出口温度增加对产成品的硫含量是正向作用，对辛烷值是负向作用；

从图 12-11 可以看出，反应器入口温度增加使得产成品的辛烷值和硫含量均增加；

从图 12-12 可以看出，D104 压力增加对产成品辛烷值表现出负向抑制作用，对硫含量表现出正向促进作用；

从图 12-13 可以看，出再生器温度增加使得产成品硫含量降低，对辛烷值影响较小；

从图 12-14 可以看出，E-101D 壳程出口管温度增加使得产成品的辛烷值和硫含量均增加；

从图 12-15 可以看出，E-101C 管程出口管温度增加使得产成品的辛烷值和硫含量均减少；

从图 12-16 可以看出，D-124 液位使得产成品硫含量增加，对辛烷值有较小的影响；

从图 12-17 可以看出，D-122 液位对产成品辛烷值影响呈现先增加后保持稳定的趋势，对产成品硫含量影响呈现先增加后减少的趋势；

从图 12-18 可以看出，再生器下部温度对产成品硫含量影响是先微增后保持平稳的趋势，对辛烷值影响是微增的趋势；

从图 12-19 可以看出，再生器下部温度对产成品的辛烷值表现出负向影响，对硫含量表现出正向影响；

从图 12-20 可以看出，D-113 压力增加对产成品的辛烷值表现出负向影响，对硫含量表现出正向影响；

从图 12-21 可以看出，低压热氮气压力对产成品的硫含量呈现出较弱的负向影响，对辛烷值是较弱的正向影响；

从图 12-22 可以看出，R-102 床层吸附剂料位密度增加对产成品辛烷值是正向影响，对硫含量是负向影响；

从图 12-23 可以看出，R-102 #3 通风挡板温度增加对产成品的硫含量影响呈现先下降后上升的趋势，对硫含量表现出正向影响；

从图 12-24 可以看出，K-102A 进气温度增加对产成品的辛烷值和硫含量均是正向影响；

从图 12-25 可以看出，F-101 出口总管压力增加对产成品硫含量影响较小，对辛烷值表现出正向影响；

从图 12-26 可以看出，F-101 出口支管#4 温度增加对产成品辛烷值影响较小，对硫含量是负向影响；

从图 12-27 可以看出，F-101 出口支管#3 温度增加对产成品辛烷值呈现较大幅度的正向影响，对硫含量影响表现出先下降后上升的趋势；

从图 12-28 可以看出，F-101 出口支管#1 温度增加对产成品的硫含量表现出中大幅度的

负向影响，对辛烷值含量表现出中大幅度的正向影响。

## 三、模型评价

### 3.1 模型优点

问题一数据处理，本研究设计了数据清理流程。首先通过最大最小幅度法将异常值数据赋值为 0，然后基于拉依达准则进一步筛选异常值并赋值为 0；将每个时间段采集的数据样本超过 20 个异常值的样本删除，剩余样本的异常值按照给定法则补齐。本研究的数据处理方法操通过两次异常值筛选，并补齐，操作流程简洁明了、高效，可快速处理大量数据，特别是在数据量庞杂时具有较好的适用性；

问题二指标筛选，构建了灰色关联模型，验证自变量与各因变量的关联度，通过排序确定关键影响因素指标。该方法克服了可线性相关以及贫信息的弊端，在非线性关系中具有较好的适用性；

问题三辛烷值损失预测模型，通过构建的 BP 神经网络模型拟合产成品辛烷值含量和硫含量与筛选出的 30 个指标的关系。构建的神经网络模型通过验证具有较好的结果。主要原因是汽油提炼各工序组成了一个非线性系统，而神经网络模型在非线性的拟合能力上具有较大优势；

问题四辛烷值损失优化模型，本研究设计了单目标优化模型并运用遗传算法对模型进行求解，在算法设计时将构建的神经网络模型嵌入遗传算法中。通过计算，可快速在 28 个变量中搜索到方程解集以满足辛烷值损失最少优化目标。通过验证该方法求解速度较快，适应多未知变量优化模型；

问题五模型可视化展示，本研究分别调整操作变量，带入辛烷值损失神经网络模型，获取不同操作变量变化对产成品辛烷值和含硫量的影响关系，并进行可视化展现，可直观观测到各变量变化对产成品辛烷值含量和硫含量的影响。

### 3.2 模型缺点

问题一数据处理，研究设置了每个样本变量数据异常值超过 20 个，则该行样本数据删除。此部分在设置异常值数量上需进一步讨论研究；

问题二主要变量筛选，在筛选时是基于数据挖掘方式对两变量间关系进行计算，属于定量分析。为考虑实际生产中各位点与产品之间的关键关系，后续研究应采用定量与定性研究结合的方式进行关键变量筛选；

问题三辛烷值损失预测模型构建，本研究构建了三层神经网络模型，只是对隐含层节点数量和相关计算参数进行讨论，但未涉及到多个隐含层测试，这也是日后深入研究的方向；

问题四辛烷值损失优化模型，该模型通过主要是通过遗传算法对模型进行求解，但是遗传算法使用于 NP 难问题很难找到最优解，只能求得较优解；

问题五数据可视化展示，该部分只展示了每个操作变量单一变化对产成品硫含量和辛烷值的影响，当多个变量同时变化时未讨论。

## 参考文献

- [1]中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 车用汽油: GB 17930—2016[S]. 北京: 中国标准出版社, 2016.
- [2]胥红玉. 浅谈汽油辛烷值的影响因素[J]. 品牌与标准化, 2020(05): 49-50+52.
- [3]朱云霞, 徐惠. S-Zorb 技术的完善及发展[J]. 炼油技术与工程, 2009, 39(08): 7-12.
- [4]刘永才, 李佳. 影响 S Zorb 装置汽油辛烷值损失因素分析[J]. 石油化工设计, 2019, 36(04): 12-15+5.
- [5] Deng J L. Introduction to Grey system theory[M]. Sci-Tech Information Services, 1989.
- [6] 罗均, 王琦. 产品设计质量灰色系统综合评价方法的研究[J]. 机械科学与技术, 2000, 19(5): 747-749.
- [7] Geem, Z.W., Transport energy demand modeling of South Korea using artificial neural network. Energy Policy, 2011. 39(8): p. 4644-4650.
- [8] 王珏, 鲍勤. 基于小波神经网络的中国能源需求预测模型[J]. 系统科学与数学, 2009, 29(11): 1542-1551.
- [9] 吕彬, 曾洁. 基于 BP 神经网络的电力负荷预测研究[J]. 现代商贸工业, 2009, 21(21): 254-255.
- [10] Sözen A, Gülseven Z, Arcaklioğlu E. Forecasting based on sectoral energy consumption of GHGs in Turkey and mitigation policies[J]. Energy Policy, 2007, 35(12): 6491-6505.
- [11] Kankal, M., et al., Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. Applied Energy, 2011. 88(5): p. 1927-1939.
- [12] 陶阳威, 孙梅, 王小芳. 基于改进的 BP 神经网络的中国能源需求预测研究[J]. 山西财经大学学报, 2010(s2): 3-5.
- [13] 张俊深, 袁程炜. 基于 BP 神经网络与修正 GM(1,1)模型的能源消费组合预测[J]. 统计与决策, 2016(5): 90-93.
- [14] Cellura M, Brano V L, Marvuglia A. Forecasting daily urban electric load profiles using artificial neural networks[J]. Energy Conversion & Management, 2004, 45(18): 2879-2900.
- [15] PETALAS Y G, PARSOPOULOS K E, VRAHATIS M N. Improving fuzzy cognitive maps learning through memetic particle swarm optimization[J]. Soft Computing, 2009, 13(1): 77-94.
- [16] GHAZANFARI M, ALIZADEH S, FATHIAN M, et al. Comparing simulated annealing and genetic algorithm in learning FCM[J]. Applied Mathematics and Computation, 2007, 192(1): 56-68.
- [17] LIU G, YANG Z, CHEN B, et al. Monitoring trends of urban development and environmental impact of Beijing, 1999–2006[J]. Science of The Total Environment, 2011, 409(18): 3295-3308.

## 附录 Matlab 源程序

针对本题所有的算法均是通过 MATLAB 软件实现，各问题求解程序如下。各问题程序源文件（.m 文件）见附件 2。程序运行中所涉及的数据表格（Excel 表格）见附件 3。

### 问题一算法程序

```
clc
clear all
%% 1、数据读取
[num,txt,row]=xlsread('data1');
[num_fanwei,txt_fanwei,row_fanwei]=xlsread('fanwei345');%读取附件中各操作变量范围
[row,col] = size(num);
fanwei = num_fanwei';
D1 = num(2:row,17:col);%获取附件 1 中所有的操作变量数据
cmax = fanwei(2,:);%找到附件 4 操作变量中最大值
cmin = fanwei(1,:);%找到附件 4 操作变量中最小值
[num285,txt285,row285] = xlsread('data285');
[num313,txt313,row313] = xlsread('data313');
D285 = num285;%获取附件三中 285 操作变量数据集
D313 = num313;%获取附件三中 313 操作变量数据集
[row285,col285] = size(D285);
[row313,col313] = size(D313);
%% 2、最大最小幅度异常值处理
D285_maxmin = zeros(row285,col285);
%对 285 操作变量按照最大最小幅度处理，超过的数据赋值 0
for j = 1:col285
    for i = 1:row285
        if D285(i,j)>cmax(j) || D285(i,j)<cmin(j)
            D285_maxmin(i,j) = 0;
        else
            D285_maxmin(i,j) = D285(i,j);
        end
    end
end
D313_maxmin = zeros(row313,col313);
%对 313 操作变量按照最大最小幅度处理，超过的数据赋值 0
for j = 1:col313
    for i = 1:row313
        if D313(i,j)>cmax(j) || D313(i,j)<cmin(j)
            D313_maxmin(i,j) = 0;
        else
            D313_maxmin(i,j) = D313(i,j);
        end
    end
end
```



```

end
zeros_num285_2 = zeros(row285,1);
for i = 1:row285
    zeros_num285_2(i) = length(find(D285_maxmin(i,:)==0)) ;
end
zeros_num313_2 = zeros(row313,1);
for i = 1:row313
    zeros_num313_2(i) = length(find(D313_maxmin(i,:)==0)) ;
end
%% 3、拉依达准则去除异常值
LD285 = zeros(row285,col285);
D285_maxmin_mean = mean(D285_maxmin); %计算各列算数平均值
D285_vi = D285_maxmin - D285_maxmin_mean;
D285_vi_pingfa = D285_vi.* D285_vi;
sum_D285_vi_pingfa = sum(D285_vi_pingfa);
dred285 =sqrt((sum_D285_vi_pingfa /(row285-1)));
for j =1:col285
    for i =1:row285
        if abs(D285_vi(i,j))>3*dred285(j)
            LD285(i,j) = 0;
        else
            LD285(i,j)=D285_maxmin(i,j);
        end
    end
end
LD313 = zeros(row313,col313);
D313_maxmin_mean = mean(D313_maxmin); %计算各列算数平均值
D313_vi = D313_maxmin - D313_maxmin_mean;
D313_vi_pingfa = D313_vi.* D313_vi;
sum_D313_vi_pingfa = sum(D313_vi_pingfa);
dred313 =sqrt((sum_D313_vi_pingfa /(row313-1)));
for j =1:col313
    for i =1:row313
        if abs(D313_vi(i,j))>3*dred313(j)
            LD313(i,j) = 0;
        else
            LD313(i,j)=D313_maxmin(i,j);
        end
    end
end
zeros_num285_3 = zeros(row285,1);
for i = 1:row285
    zeros_num285_3(i) = length(find(LD285(i,:)==0)) ;
end

```

```

zeros_num313_3 = zeros(row313,1);
for i = 1:row313
    zeros_num313_3(i) = length(find(LD313(i,:)==0)) ;
end
%% 4、找到每行大于 0 个数大于 30 的行号并剔除
hanghao285 = find(zeros_num285_3>20);
[row_hanghao285,col_hanghao285] = size(hanghao285);
LD285(hanghao285,:) = [];
quling_LD285 = LD285;
hanghao313 = find(zeros_num313_3>20);
[row_hanghao313,col_hanghao313] = size(hanghao313);
LD313(hanghao313,:) = [];
quling_LD313 = LD313;
%% 5、数据补齐
[row284_last,col285_last] = size(quling_LD285);
[row313_last,col313_last] = size(quling_LD313);
twohours_mean285 = mean(quling_LD285);
twohours_mean313 = mean(quling_LD313);
last_285 = zeros(row284_last,col285_last);
last_313 = zeros(row313_last,col313_last);
for i = 1:row284_last
    for j = 1:col285_last
        if quling_LD285(i,j) == 0
            last_285(i,j) = twohours_mean285 (1,j);
        else
            last_285(i,j) = quling_LD285(i,j);
        end
    end
end
for i = 1:row313_last
    for j = 1:col313_last
        if quling_LD313(i,j) == 0
            last_313(i,j) = twohours_mean313(1,j);
        else
            last_313(i,j) = quling_LD313(i,j);
        end
    end
end
%% 5、求均值
mean_last285 = mean(last_285);
mean_last313 = mean(last_313);
问题二算法程序

clc
clear all

```

```

[num,txt,row]=xlsread('q2data1');
x = num(2:end,2:end)';
index_num = size(x,1)
column_num = size(x,2);
% 1、数据均值化处理
x_mean = mean(x,2);
for i = 1:index_num
    x(i,:) = x(i,+)/x_mean(i,1);
end
%2、提取参考列和比较列数据
ck = x(1,:);
cp=x(2:end,:);
cp_index_num = size(cp,1);
y = cp;
x = ck;
y_row = size(y,1);%%计算矩阵 y 的行数
y_col =size(y,2);%%计算矩阵 y 的列数
x_col = size(x,2);%%计算 x 的列数
if y_col ~= x_col
    error(message('MATLAB:greyrelation:wrong in input data'));
end
temp_y = y;%%绝对关联度中比较序列中的数据处理后的矩阵
temp_x = x;%%x 数据处理后的矩阵
for i =1:x_col
    temp = x(i)-x(1);
    temp_x(i)=temp;
end
for i =1:y_row
    for j=1:y_col
        temp = y(i,j) - y(i,1);
        temp_y(i,j)=temp;
    end
end%处理过程
%temp_x;
%temp_y;
s0 = abs(sum(temp_x)-0.5*temp_x(x_col));
abs_xy =[];
for i=1:y_row
    si = abs(sum(temp_y(i,:))-0.5*temp_y(i,y_col));
    si_s0 = abs(si-s0);
    abs_xy(i,1) =(1+s0+si)/(1+s0+si+si_s0);
end
%下面开始计算相关关联度
temp_y2 = y;

```

```

temp_x2 = x;
for i = 1:x_col
    temp = x(i)/x(1);
    temp_x2(i)=temp-1;
end
for i = 1:y_row
    for j = 1:y_col
        temp = y(i,j) / y(i,1);
        temp_y2(i,j)=temp-1;
    end
end
s02 = abs(sum(temp_x2)-0.5*temp_x2(x_col));
rela_xy=[];
for i = 1:y_row
    si2 = abs(sum(temp_y2(i,:))-0.5*temp_y2(i,y_col));
    si2_s02 = abs(si2-s02);
    rela_xy(i,1) =(1+s02+si2)/(1+s02+si2+si2_s02)
end
%下面计算综合关联度
com_xy = 0.5*abs_xy +(1-0.5)*rela_xy;%返回的是综合关联度

```

### 问题三算法程序

```

clc
clear
close all
[num_input,txt_input,raw_input]=xlsread('input') ;
[num_output,txt_output,raw_output]=xlsread('output') ;
input_0 = num_input(:,3:end);
input = num_input;
output = num_output;
data_train_input = input(1:300,:);
data_train_output = output(1:300,:);
data_test_input = input(301:end,:);
data_test_output = output(301:end,:);
input_train = data_train_input';
output_train = data_train_output';
input_test = data_test_input';
output_test = data_test_output';
%训练数据归一化
[inputn,inputps] = mapminmax(input_train);
[outputn,outputps] = mapminmax(output_train);
net = newff(inputn,outputn,90);
%参数设置
net.trainParam.epochs=100;%迭代次数

```

```

net.trainParam.lr=0.4;%学习率
net.trainParam.goal=0.0000000001;%收敛目标
%神经网络训练
net = train(net,inputn,outputn);
%训练数据归一化
inputn_test = mapminmax('apply',input_test,inputps);
%神经网络测试输出
an = sim(net,inputn_test);
BPoutput = mapminmax('reverse',an,outputps);
%数据可视化
figure(1)
plot(BPoutput(1,:), 'r') %红
hold on
plot(output_test(1,:), 'b');
hold on
plot(BPoutput(2,:), 'k') %
hold on
plot(output_test(2,:), 'g');
legend('模拟值(含硫量)', '原始值(含硫量)', '模拟值(辛烷值)', '原始值(辛烷值)')
err = abs(BPoutput - output_test);
err_mean = mean(err);
figure(2)
plot(err_mean, '-*')
title('测试误差')
ylabel('平均误差')
xlabel('样本')

```

## 问题四算法程序

```

clc
clear
close all
x1 = 89.8; %设定原材料辛烷值
x2 = 56.10; %饱和烃, v% (烷烃+环烷烃)
x_index = 173; %优化数据编码
y0 = 89.22; %原始数据产成品辛烷值
[num_input, txt_input, raw_input] = xlsread('input');
[num_output, txt_output, raw_output] = xlsread('output');
input = num_input;
output = num_output;
data_train_input = input(1:300,:);
data_train_output = output(1:300,:);
input_train = data_train_input'; %30*300
output_train = data_train_output'; %2*300
%训练数据归一化

```

```

[inputn,inputps] = mapminmax(input_train);
[outputn,outputps] = mapminmax(output_train);
%参数设置
net = newff(inputn,outputn,90);
net.trainParam.epochs=100;%迭代次数
net.trainParam.lr=0.4;%学习率
net.trainParam.goal=0.0000000001;%收敛目标
%神经网络训练
net = train(net,inputn,outputn);
%对 x1, x2 标准化数据获取
x11 = inputn(1,x_index);
x22 = inputn(2,x_index);
%% 遗传参数设置
popsize = 50; %种群大小
pc = 0.7; %交叉率
pm = 0.09; %变异率
Iteration = 1000; %最大迭代次数
nodes = 28;
r_lost_mean = 1.2; %辛烷值损失值均值
%% 初始化总群
initPop = zeros(nodes,popsize);
for j = 1:popsize
    r = rand(nodes+2,popsize);
    for i = 1:nodes+2
        initPop(i,j) = -1+2*r(i,j);
    end
    r=[];
end
initPop(1,:)=x11;
initPop(2,:)=x22;
trace =zeros(Iteration,3);%第一列存迭代次数，第二列存最小误差，第三列存平均误差
NewPop = zeros(nodes+2,popsize);%选择后的种群
for gen = 1:Iteration
    %% 选择操作
    % 计算适应度
    an = sim(net,initPop);
    an_output = mapminmax('reverse',an,outputps);
    fitness = zeros(1,popsize);
    for j = 1:popsize
        if an_output(1,j)<0||((x1-an_output(2,j)-r_lost_mean)/r_lost_mean>0.3||((x1-
an_output(2,j))<0||an_output(1,j)>5
            fitness(j) = 0.000001;
        else
            fitness(j) = 1/(x1-an_output(2,j));
        end
    end
end

```

```

        end
    end
[mmin index_err] = max(fitness);
    %计算选择概率
    pz = fitness./sum(fitness);
    %计算概率累计
    qz = sum(pz);
    %执行选择
    index = zeros(1,popsize);
    for i = 1:popsize
        pick = rand;
        while pick == 0
            pick = rand;
        end
        pick = pick - pz(index_err);
        if pick<0
            index(1,i) = index_err;
        else
            index(1,i) = ceil(popsize*rand);
        end
    end
    end
NewPop = initPop(:,index);
%% 交叉操作
for j =1:popsize
    pick = rand(1,2);
    while prod(pick) == 0
        pick = rand(1,2);
    end
    index = ceil(pick * popsize);
    %交叉率是否决定交叉
    pick = rand;
    while pick == 0
        pick = rand;
    end
    if pick > pc
        continue
    end
    flag = 0;
    pick1 = rand;
    pick2 = rand;
    pos1 = ceil(pick1 * (nodes+2));
    pos2 = ceil(pick2 * (nodes+2));
    while pos1 == pos2
        pick1 = rand;

```

```

        pick2 = rand;
        pos1 = ceil(pick1 * nodes);
        pos2 = ceil(pick2 * nodes);
        pos1 = min(pos1,pos2);
        pos2 = max(pos1,pos2);
        v1 = NewPop(pos1:pos2,index(1));
        v2 = NewPop(pos1:pos2,index(2));
        NewPop(pos1:pos2,index(1)) = v2;
        NewPop(pos1:pos2,index(2)) = v1;
    end
end

%% 变异操作
for j =1:popsiz
    r1 = rand(nodes+2,popsiz);
    pick = rand;
    if pick > pm
        continue;
    end
    %变异位置
    pick = rand;
    while pick == 0
        pick = rand;
    end
    index = ceil(pick*(nodes+2));
    pickj = rand;
    while pickj == 0
        pickj = rand;
    end
    j=ceil(pickj*(nodes+2));
    if index==1
        NewPop(index,j) = x11;
    elseif index==2
        NewPop(index,j) = x22;
    else
        NewPop(index,j) = -1+2*r1(index,j);
    end
end
end
%计算误差
bn = sim(net,NewPop);
bn_output = mapminmax('reverse',bn,outputps);
err1 =zeros(1,popsiz);
for i =1:popsiz
    err1(1,i) = abs(x1-bn_output(2,i));
end

```



```

end
%计算每代平均误差
aveErr = sum(err1)/popsize;
[minErr,bestIndex] = min(err1);
bestinputps = NewPop(:,bestIndex);
trace (gen,1) = gen;
trace (gen,2) = minErr;
trace (gen,3) = aveErr;
initPop = NewPop;
end
x = trace(:,1);
minerr = trace(:,2);
avgerr = trace(:,3);
figure
plot(x,minerr,'r--',x,avgerr,'b-');
xlabel('Iterations');
ylabel('ERR');
legend('minerr','avgerr');
grid;
an = sim(net,bestinputps);
BPoutput = mapminmax('reverse',an,outputps)
bsetChrombsetChrom_rever= mapminmax('reverse',bestinputps,inputps);

```

### 问题五算法程序

```

clc
clear
close all
[num_input,txt_input,raw_input]=xlsread('input') ;
[num_output,txt_output,raw_output]=xlsread('output') ;
input_0 = num_input(:,3:end);
input = num_input;
output = num_output;
data_train_input = input(1:300,:);
data_train_output = output(1:300,:);
input_train = data_train_input';
output_train = data_train_output';
[num_fanwei,txt_input,raw_input]=xlsread('fanwei5') ;
[num_derd,txt_derd,raw_derd]=xlsread('derd') ;
fenwei_min = num_fanwei(:,1);
fenwei_max = num_fanwei(:,2);
index = 30;
num_index = fenwei_min(index):num_derd:fenwei_max(index)
[index_row index_l] = size(num_index)
input_test = zeros(30,index_l);
for i=1:30

```

```

        for j=1:index_l
            if i==index
                input_test(i,j)=num_index (1,j);
            else
                input_test(i,j)=fenwei_min(i,1);
            end
        end
    end
end
%训练数据归一化
[inputn,inputps] = mapminmax(input_train);
[outputn,outputps] = mapminmax(output_train);
net = newff(inputn,outputn,90);
%参数设置
net.trainParam.epochs=10000;%迭代次数
net.trainParam.lr=0.4;%学习率
net.trainParam.goal=0.0000000001;%收敛目标
%神经网络训练
net = train(net,inputn,outputn);
%训练数据归一化
inputn_test = mapminmax('apply',input_test,inputps);
%神经网络测试输出
an = sim(net,inputn_test);
BPoutput = mapminmax('reverse',an,outputps);
%数据可视化
figure(1)
plot(BPoutput(1,:), 'r') %红
hold on
plot(BPoutput(2,:), 'b') %
legend('（硫含量）','辛烷值','fontsize',14)
xlabel('步长')

```